



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 212039**

**TO: Manjunath N Rao**  
**Location: REM/2A01/2C70**  
**Art Unit: 1652**  
**Monday, January 08, 2007**  
**Case Serial Number: 10/800684**

**From: Kristine Hensle**  
**Location: Biotech-Chem Library**  
**REM-1A70**  
**Phone: (571)272-4161**

**Kristine.Hensle@uspto.gov**

### **Search Notes**

Examiner Rao,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Kristine Hensle  
Reference Librarian  
STIC Biotech/Chem Library  
(571)272-4161

**This Page Blank (uspto)**

**STIC-Biotech/ChemLib**

---

**From:** Chan, Christina  
**Sent:** Friday, January 05, 2007 4:43 PM  
**To:** Rao, Manjunath N.; STIC-Biotech/ChemLib  
**Subject:** RE: Request for RUSH sequence search for 10/800684

Please rush. Thanks Chris

Chris Chan  
TC 1600 New Hire Training Coordinator and SPE 1644  
(571)-272-0841  
Remsen, 3E89

-----Original Message-----

**From:** Rao, Manjunath N.  
**Sent:** Friday, January 05, 2007 2:48 PM  
**To:** Chan, Christina  
**Subject:** Request for RUSH sequence search for 10/800684

Hello Christina,

Please authorize the request below as RUSH. The reason being this is an amended case that is due in the next bi-week.

Thanks  
-Manjunath

*Manjunath N. Rao, Ph.D.*  
Primary Patent Examiner  
Art Unit 1652, Room 2A01  
Remsen Building  
United States Patent and Trademark Office  
400, Dulany St, VA

Phone: 571-272-0939  
Fax: 571-273-0939

**From: Manjunath N. Rao**  
**Art Unit 1652, Room 2A01**  
**Mail Box in Room 2C70**  
Phone: 272-0939

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date completed: \_\_\_\_\_  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# \_\_\_\_\_ AA# \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

**This Page Blank (uspto)**



Date: 1-5-07

Please search the following as soon as possible for application with serial number

**10/800684**

1. SEQ ID NO: 1, and 5 against all commercial amino acid sequence databases, issued patents/published applications amino acid sequence database and pending application amino acid sequence database. Please provide a print of results.

If you have any questions please call me at the above phone number.

Thanks

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date completed: \_\_\_\_\_  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# \_\_\_\_\_ AA# \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

**This Page Blank (uspto)**

# STIC SEARCH RESULTS FEEDBACK FORM

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact **the searcher or contact:**

Mary Hale, Information Branch Supervisor  
571-272-2507 Remsen 1 A51

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg.

**This Page Blank (uspto)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2007, 22:19:02 ; Search time 142 Seconds  
(without alignments)  
2121.871 Million cell updates/sec

Title: US-10-800-684-1  
Perfect score: 3437  
Sequence: 1 MKRLCAVVALVVLVGLAGT.....YAYSTYGVADYQLKAVVYVG 659

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq.8:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*  
9: geneseqp2005s:\*  
10: geneseqp2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match % | Length | DB ID | Description        |
|------------|--------|---------------|--------|-------|--------------------|
| 1          | 3437   | 100.0         | 659    | 2     | AAW24121 Thermococ |
| 2          | 3437   | 100.0         | 659    | 2     | AAW94840 W09856926 |
| 3          | 2914   | 84.8          | 659    | 2     | AAW24123 Protease. |
| 4          | 2664   | 77.5          | 663    | 8     | ADN46373 Thermococ |
| 5          | 2483.5 | 72.3          | 654    | 2     | AAW24129 Pyrococcu |
| 6          | 2483.5 | 72.3          | 654    | 2     | AAW94841 Hyperther |
| 7          | 2138.5 | 62.2          | 522    | 2     | AAW24122 Pyrococcu |
| 8          | 2138.5 | 62.2          | 522    | 2     | AAW94838 Hyperther |
| 9          | 1707   | 49.7          | 412    | 2     | AAW94836 Hyperther |
| 10         | 1154   | 33.6          | 237    | 2     | AAW87009 Hyperther |
| 11         | 935    | 27.2          | 188    | 2     | AAW87011 Peptide s |
| 12         | 935    | 27.2          | 188    | 2     | AAW24126 Thermococ |
| 13         | 918.5  | 26.7          | 545    | 4     | ABW09483 T. yonsei |
| 14         | 665    | 19.3          | 734    | 4     | ABW09483 Streptomy |
| 15         | 665    | 19.3          | 823    | 2     | AAW13667 DHPA-mel  |
| 16         | 647.5  | 18.8          | 520    | 2     | AAW13668 Fragment  |
| 17         | 636    | 18.5          | 1079   | 4     | AAW81180 Transglut |
| 18         | 636    | 18.5          | 1079   | 6     | ABW07391 Foreign p |
| 19         | 632.5  | 18.4          | 443    | 8     | ADM99125 Bacterial |
| 20         | 632.5  | 18.4          | 1237   | 8     | ABW11343 Protein e |
| 21         | 590.5  | 17.9          | 1227   | 8     | ADM99123 Bacterial |
| 22         | 590.5  | 17.2          | 182    | 4     | ABW09484 T. yonsei |
| 23         | 577.5  | 16.8          | 168    | 2     | AAW24127 Pyrococcu |

|    |       |      |     |    |                    |
|----|-------|------|-----|----|--------------------|
| 24 | 576.5 | 16.8 | 552 | 8  | ADM99241 Environme |
| 25 | 565.5 | 16.5 | 383 | 5  | ADW54158 L-25 Kera |
| 26 | 564.5 | 16.4 | 379 | 6  | AAE29944 Bacillus  |
| 27 | 563.5 | 16.4 | 379 | 10 | AEF11989 Bacillus  |
| 28 | 561.5 | 16.3 | 379 | 2  | AAW4137 Sequence   |
| 29 | 560.5 | 16.3 | 379 | 2  | AAW88521 Keratins  |
| 30 | 559.5 | 16.3 | 379 | 5  | AAW26762 Bacillus  |
| 31 | 559.5 | 16.3 | 379 | 2  | AAE19062 Bacillus  |
| 32 | 557.5 | 16.2 | 350 | 2  | AAW46598 Amino aci |
| 33 | 557.5 | 16.2 | 379 | 2  | AAW1600 Subtilisi  |
| 34 | 557.5 | 16.2 | 379 | 6  | AAE29943 Bacillus  |
| 35 | 557.5 | 16.2 | 379 | 7  | ADW70896 Subtilisi |
| 36 | 557.5 | 16.2 | 379 | 8  | ADW39841 B. amylo  |
| 37 | 557.5 | 16.2 | 379 | 8  | ADW58783 B. lichen |
| 38 | 557.5 | 16.2 | 379 | 8  | ADW64368 Bacillus  |
| 39 | 557.5 | 16.2 | 380 | 8  | ADW39843 Subtilisi |
| 40 | 556.5 | 16.2 | 379 | 6  | AAE29945 Bacillus  |
| 41 | 555.5 | 16.2 | 379 | 8  | ADW58791 B. lichen |
| 42 | 554.5 | 16.1 | 372 | 6  | AAE29942 Bacillus  |
| 43 | 554.5 | 16.1 | 379 | 8  | ADW58786 B. lichen |
| 44 | 551.5 | 16.0 | 378 | 6  | AAE29947 Bacillus  |
| 45 | 550.5 | 16.0 | 379 | 6  | AAE29946 Bacillus  |

## ALIGNMENTS

|          |   |
|----------|---|
| RESULT 1 | AAW24121 standard; protein; 659 AA.                                       |
| AAW24121 |   |
| XX       | AAW24121;   |
| AC       | 17-OCT-2003 (revised)   |
| XX       | 20-APR-1998 (first entry)   |
| DT       | Thermococcus protease.  |
| XX       | Protease; research reagent; thermal stability; thermococcus celer.        |
| XX       | Thermococcus celer; DSM-2476.   |
| OS       | WO9721823-A1.   |
| XX       | 19-JUN-1997.  |
| XX       | 07-NOV-1996; 96WO-JP003253.   |
| XX       | 12-DEC-1995; 95JP-00323285.   |
| XX       | (TAKI ) TAKARA SHUZO CO LTD.  |
| PA       | Takakura H, Morishita M, Yamamoto K, Mitra M, Asada K;                    |
| XX       | Tsunasawa S, Kato I;  |
| PI       | WPI; 1997-332794/30.  |
| XX       | N-PSDB; AAT85667.   |
| DR       | Protease(s) and genes encoding them obtained from Thermococcus and        |
| XX       | Pyrococcus strains - have extremely high thermal stability and are useful |
| PT       | industrially and as research reagents.                                    |
| XX       | Claim 1; Page 82-85; 159pp; Japanese.                                     |
| PS       | This sequence represents the protease from Thermococcus celer DSM-2476.   |
| XX       | This sequence is a protease of the invention. The proteases of the        |
| CC       | invention have extremely high thermal stability. The proteases can be     |
| CC       | used as research reagents, and industrially in the food, drug and         |
| CC       | chemical industries. (updated on 17-OCT-2003 to standardise OS field)     |
| XX       | Sequence 659 AA;  |
| SQ       | Query Match 100.0%; Score 3437; DB 2; Length 659;                         |

Best Local Similarity 100.0%; Pred. No. 5.7e-203;  
Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MKRLGAVVIALVVLGILAGTALAAPVKPYVRNNAAVOOKYGGTLTGELFKKQVRMMNNEV 60
DB 1 MKRLGAVVIALVVLGILAGTALAAPVKPYVRNNAAVOOKYGGTLTGELFKKQVRMMNNEV 60
QY 61 DTVYMGSGYGDRAVKVRLMGAQVYSYKIIIPAAVKIKARDLLIAGMIDTGYFGNT 120
DB 61 DTVYMGSGYGDRAVKVRLMGAQVYSYKIIIPAAVKIKARDLLIAGMIDTGYFGNT 120
QY 121 RVSGIKFIEDYKVOVDATSVSQIGADTVNNSLGYSQGVVVALVDTGIDANHPDLKXK 180
DB 121 RVSGIKFIEDYKVOVDATSVSQIGADTVNNSLGYSQGVVVALVDTGIDANHPDLKXK 180
QY 181 VIGWYDAVNGRSTPYDDQGHGTHVAGIAGTGSVNSQYIIGVAPAKLVGVVVLGADSGS 240
DB 181 VIGWYDAVNGRSTPYDDQGHGTHVAGIAGTGSVNSQYIIGVAPAKLVGVVVLGADSGS 240
QY 241 VSTIIAGVDMVWVQNDKXGIRVINISLSSSQSSDGTDSLQAANNAMDAIIVCVAAGNS 300
DB 241 VSTIIAGVDMVWVQNDKXGIRVINISLSSSQSSDGTDSLQAANNAMDAIIVCVAAGNS 300
QY 301 GPNITVGSPPAAASKVITVGAVDSDNDIASFSSRGPTADGRLKBEVAPGVDTIAPRASG 360
DB 301 GPNITVGSPPAAASKVITVGAVDSDNDIASFSSRGPTADGRLKBEVAPGVDTIAPRASG 360
QY 361 TSMGTPINDYTTKASGTSMTPHVSGVALILQAHPSWTPDKVKTALIEADIAAPEIA 420
DB 361 TSMGTPINDYTTKASGTSMTPHVSGVALILQAHPSWTPDKVKTALIEADIAAPEIA 420
QY 421 DIAGAGRVNYYKAIKXDDYAKLFTFGSVADKSGATFTFYSGATFTATILYMDTSSDI 480
DB 421 DIAGAGRVNYYKAIKXDDYAKLFTFGSVADKSGATFTFYSGATFTATILYMDTSSDI 480
QY 481 DLYLYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTKVVSYKGAANYQVDVVSDDSLQ 540
DB 481 DLYLYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTKVVSYKGAANYQVDVVSDDSLQ 540
QY 541 SGGGNPNPNPNPTPTTDTOTFTGSVNDYWDTSDFTTMNVNSGATKITGDLTFDTSYND 600
DB 541 SGGGNPNPNPNPTPTTDTOTFTGSVNDYWDTSDFTTMNVNSGATKITGDLTFDTSYND 600
QY 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVLYASTYGADYQLAAVYYG 659
DB 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVLYASTYGADYQLAAVYYG 659

```

## RESULT 2

AAW94840 standard; protein; 659 AA.

```

XX ID AAW94840 standard; protein; 659 AA.
XX AC AAW94840;
XX DT 06-MAY-1999 (first entry)
XX DE W09856926 Seq ID 12.
XX OS Thermococcus celer.
XX PN W09856926-A1.
XX PD 17-DEC-1998.
XX PF 04-JUN-1998; 98MO-JP002465.
XX PR 10-JUN-1997; 97JP-00151693.
XX PA (TAKI) TAKARA SHUZO CO LTD.
XX

```

PI Takakura H, Morishita M, Shimoto T, Asada K, Kato I;  
XX WPI: 1999-080907/07.  
XX N-PSDB; AAX05926.  
PT Recombinant hyperthermostable protease from *Pyrococcus furiosus* - and  
PT gene encoding it, for large scale production of the protease for  
XX industrial use.  
XX Disclosure; Page 55-58; 82pp; Japanese.  
XX  
XX The invention relates to a hyperthermostable protease derived from a  
XX thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has  
XX working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10  
XX (optimum 6-8), and retains more than 90% of its activity after 8 hours at  
XX 95 deg.C. The invention also provides gene sequences encoding a  
XX polypeptide of formula SIG-Ala-Gly-Gly-Asn-Pro, where SIG is a signal  
XX peptide from *Bacillus* strains, and PRO is the above protease. Host cells  
XX (especially *Bacillus* strains) transformed with vectors comprising the  
XX genes are used for the recombinant production of the protease. The  
XX hyperthermostable protease which can be prepared in quantity suitable for  
XX industrial use, can be used as an additive for drugs, washing agents and  
XX foodstuffs and for chemical synthesis

Sequence 659 AA:

Query Match 100.0%; Score 3437; DB 2; Length 659;  
Best Local Similarity 100.0%; Pred. No. 5.7e-203;  
Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MKRLGAVVIALVVLGILAGTALAAPVKPYVRNNAAVOOKYGGTLTGELFKKQVRMMNNEV 60
DB 1 MKRLGAVVIALVVLGILAGTALAAPVKPYVRNNAAVOOKYGGTLTGELFKKQVRMMNNEV 60
QY 61 DTVYMGSGYGDRAVKVRLMGAQVYSYKIIIPAAVKIKARDLLIAGMIDTGYFGNT 120
DB 61 DTVYMGSGYGDRAVKVRLMGAQVYSYKIIIPAAVKIKARDLLIAGMIDTGYFGNT 120
QY 121 RVSGIKFIEDYKVOVDATSVSQIGADTVNNSLGYSQGVVVALVDTGIDANHPDLKXK 180
DB 121 RVSGIKFIEDYKVOVDATSVSQIGADTVNNSLGYSQGVVVALVDTGIDANHPDLKXK 180
QY 181 VIGWYDAVNGRSTPYDDQGHGTHVAGIAGTGSVNSQYIIGVAPAKLVGVVVLGADSGS 240
DB 181 VIGWYDAVNGRSTPYDDQGHGTHVAGIAGTGSVNSQYIIGVAPAKLVGVVVLGADSGS 240
QY 241 VSTIIAGVDMVWVQNDKXGIRVINISLSSSQSSDGTDSLQAANNAMDAIIVCVAAGNS 300
DB 241 VSTIIAGVDMVWVQNDKXGIRVINISLSSSQSSDGTDSLQAANNAMDAIIVCVAAGNS 300
QY 301 GPNITVGSPPAAASKVITVGAVDSDNDIASFSSRGPTADGRLKBEVAPGVDTIAPRASG 360
DB 301 GPNITVGSPPAAASKVITVGAVDSDNDIASFSSRGPTADGRLKBEVAPGVDTIAPRASG 360
QY 361 TSMGTPINDYTTKASGTSMTPHVSGVALILQAHPSWTPDKVKTALIEADIAAPEIA 420
DB 361 TSMGTPINDYTTKASGTSMTPHVSGVALILQAHPSWTPDKVKTALIEADIAAPEIA 420
QY 421 DIAGAGRVNYYKAIKXDDYAKLFTFGSVADKSGATFTFYSGATFTATILYMDTSSDI 480
DB 421 DIAGAGRVNYYKAIKXDDYAKLFTFGSVADKSGATFTFYSGATFTATILYMDTSSDI 480
QY 481 DLYLYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTKVVSYKGAANYQVDVVSDDSLQ 540
DB 481 DLYLYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTKVVSYKGAANYQVDVVSDDSLQ 540
QY 541 SGGGNPNPNPNPTPTTDTOTFTGSVNDYWDTSDFTTMNVNSGATKITGDLTFDTSYND 600
DB 541 SGGGNPNPNPNPTPTTDTOTFTGSVNDYWDTSDFTTMNVNSGATKITGDLTFDTSYND 600
QY 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVLYASTYGADYQLAAVYYG 659
DB 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVLYASTYGADYQLAAVYYG 659

```

## RESULT 3

AAW24123

ID AAW24123 standard; protein; 659 AA.

XX AAW24123;

DT 20-APR-1998 (first entry)

XX Protease.

XX Protease; research reagent; thermal stability.

XX Synthetic.

XX MO9721823-A1.

XX 19-JUN-1997.

XX 07-NOV-1996; 96WO-JP003253.

XX 12-DEC-1995; 95JP-00323285.

XX (TAKI ) TAKARA SHUZO CO LTD.

XX Takakura H, Morishita M, Yamamoto K, Mitra M, Asada K;

XX Tsunawasa S, Kato I;

XX WPI; 1997-332794/30.

XX N-PSDB; AAT85669.

PT Protease(s) and genes encoding them obtained from Thermococcus and  
 PT Pyrococcus strains - have extremely high thermal stability and are useful  
 PT industrially and as research reagents.

PT Claim 9; Page 92-95; 159pp; Japanese.

CC This sequence is a protease of the invention. The proteases of the  
 CC invention have extremely high thermal stability. The proteases can be  
 CC used as research reagents, and industrially in the food, drug and  
 CC chemical industries

SQ Sequence 659 AA;

Query Match 84.8%; Score 2914; DB 2; Length 659;

Best Local Similarity 83.7%; Pred. No. 8.5e-171;

Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps 7;

QY 1 MERLGAVLALVGLLACTALAPKPY--VRNNAVQKNGLLTPGLFKKVGQRMNMQ 58

DB 1 MKGLALILVILVGLVGSVAAPKKEQVNN--VEKNYGLLPGLFRKIQKLNPRE 57

QY 59 EVDVYIMFGSYGDRDAVAVKRLMGAQVKSYYKIIPAVAVKIKARDLLIAGMIDTGYFG 118

DB 58 EISTYIVFENHBEKEIAVAVLELMGAKVYVYHIIPALADIKRDLVYISGL--TG--G 113

QY 119 NTRVSGIKFTQEDYKQVDDA----TSVSQIGADTVNNSLGYDSGVVAIVDTGIDAN 173

DB 114 KAKLGSVRFIQEDYKQVVAELEGDESAAQVATVVMN--LGVDGSGITIGIIDGIDS 172

QY 174 HEDLKGKVGWADVAVNGRSTPYDDQGHVHVGIVAGTQSV--NSQVIGAPAKVGVGVV 232

DB 173 HEDLQKGVGWADVAVNGRSTPYDDHGHVHVASIAAGTAPASNGKYGAPAKLAGIRV 232

QY 233 LGADSGSVSTIAGVAVVONKDKYKIRVINLSLSSQSSDGTDSLGAQVNNAMDAIV 292

DB 233 LGADSGSVSTIIGVEMAVVNDKDKYKIRVINLSLSSQSSDGTDSLGAQVNNAMDAIV 292

QY 293 VCVAAAGNSGPNTYTVGSPPAAASKVITVGAVDNSNDINIASFSSRGPPTADGRLKEBVAVPGVD 352

DB 293 VCVAAAGNSGPNTYTVGSPPAAASKVITVGAVDNSNDINIASFSSRGPPTADGRLKEBVAVPGVD 352

## RESULT 4

ADN46373

ID ADN46373 standard; protein; 663 AA.

XX ADN46373;

DT 01-JUL-2004 (first entry)

DE Thermococcus kodakarensis KOD1 protein sequence SeqID251.

XX gene disruption; gene targeting; marker gene; transformation;

XX homologous recombination; hyperthermostable archaeobacterium; KOD1;

XX gene structure; gene function; enzyme activity; medicine;

XX forensic science; food; drug inspection; molecular biology; immunology.

XX Thermococcus kodakarensis.

XX WO2004022736-A1.

XX 18-MAR-2004.

XX 29-AUG-2003; 2003WO-IB003597.

XX 30-AUG-2002; 2002JP-00319011.

XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.

XX Imanaka T, Atomi H;

XX WPI; 2004-257583/24.

PT Method for disrupting targeted gene in genome of organism particularly

PT thermotable bacterium and with genome chips for analysis, applicable in

PT studying gene structure and functions.

PT Claim 9; SEQ ID NO 251; 598pp; Japanese.

CC This invention relates to a novel method for targeting disruption of an

CC arbitrary gene in a genome of an organism which comprises providing the

CC whole sequential data of the genome of such organism, selecting at least

CC 1 arbitrary region in the sequence, providing a vector that contains a

CC sequence homologous with the selected region and a marker gene,

CC transformation, and homologous recombination. The genome is preferably

CC the genome of a hyperthermostable archaeobacterium, particularly

CC Thermococcus kodakarensis KOD1. The method is for targeting the

CC disruption of a gene in the genome of an organism, which is applicable in

CC studying gene structure and functions as well as enzyme activities of

CC encoded proteins and useful in medicine, forensic science, food or drug

CC inspection, molecular biology and immunology. With this method, the  
 CC disruption of a gene at an arbitrary position in a genome can be achieved  
 CC efficiently and reliably. The present sequence is that of a protein  
 CC encoded by the genome of *Thermococcus kodakarensis* which was derived  
 CC using the method of the invention. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 663 AA;

Query Match 77.5%; Score 2664; DB 8; Length 663;

Best Local Similarity 75.9%; Pred. No. 2.1e-155;  
 Matches 505; Conservative 64; Mismatches 88; Indels 8; Gaps 4;

```

QY 1 MKRLGAVLVALLVGLGALTAALPAVKPVYRNNAVQKNGYGLTPGIFKVKQVMNNQYV 60
DB 1 MKKFGAVVLLFLVGLMAGSVLAAPQKPAVR-NVSOQKNYGLLTPGLFKVKQVMNNQYV 59
QY 61 DTIVIMFGSYGDRDAVAVVLRMLGAQVYKXKIIPAVVVKIKARDLLIAGMIDTGYRGNT 120
DB 60 STIIMFQNDKEKAVSILDFLGAKIKYNNHIIIPALAVKIKVLDLIIAGMIDTGYFGNA 119
QY 121 RVSGIKFIQEDYKYQVDDA-----SVSOIGADTVNNSLGYDGSQVVAIVDTGIDANHP 175
DB 120 QLSGVQFIQEDYKYKAVETBGLDESAQVMAITMMN-LGYDGSGITIGIIDTIDASHP 178
QY 176 DLKRGVIGWYDAVNGRSTPYDDQGHGTHVAGVAGTSV-NSQYIGVAPGAKLVGVVVG 234
DB 179 DLQGVIGWYDFVNGKTPPYDDNGHGHVSAIAGTAANSKKGKGAAPAKLVGIVLVN 238
QY 235 ADGSGYSTIIAGVWVYVONKDKGIRVYNISLGSQSSQSDGTSLSOAVNNAMDAIIVC 294
DB 239 GQSGSISDIINGVMAVONKDKGIRVYNISLGSQSSQSDGTSLSOAVNNAMDAIIVV 298
QY 295 VVAAGNSGPNYTYVGSPPAAASKVITVGAVDSDNINASFSSRGPADGR.LKEVVA PGVDI 354
DB 299 VVAAGNSGPNKYTVGSPPAAASKVITVGAVDKYDVTIDPSSRGPADNRLKPEVVA PGWMI 358
QY 355 APRAAGTSMGTPINDYTYTKASGTSMAIPIHVSQVALLIQAPHSWTPDKVKTAL.IETADIV 414
DB 359 AARASGTSMGQPINDYTYTAAPGTSMAIPIHVSQVALLIQAPHSWTPDKVKTAL.IETADIV 418
QY 415 APKEIADIAYAGRVNVYKAIKYDDYAKLFTFGSVAADGSAITHTFEDVSGATFTATLYMD 474
DB 419 KPDEIADIAYAGRVNAKAAIYDDYAKLFTFGVSNKGSQSHQPTISGAFVTAIILYMD 478
QY 475 TGSSDIDLXYLYDPNGNEVDYSYTAAYGFEKGYNNPTAGTWTVKVSYKGAANYQVDVVS 534
DB 479 NSGSDLDLYLYDPNGNQVDYSYTAAYGFEKGYNNPTAGTWTIKVSYSGSANYQVDVVS 538
QY 535 DGSLSQSGGPNPNPNPNPTPTDTQTFTFGSVDNDYMDTSPTFTMNNVNSGATKKTIGDLTF 594
DB 539 DGSLSQSGGSESPSPSPPTVDEKFTFTGIVHDYDKSDPTFTMTVNSGATKKTIGDLTF 598
QY 595 DTSVNDLILYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFLYAASTGMADYOLKA 654
DB 599 DTSVNDLILYLYDPNQNLVDRSSNSYEHVEYINNPAPGTWTFLYVYADTIGYADYOLDA 658
QY 655 VVYYG 659
DB 659 KVVYV 663

```

RESULT 5  
 AAM24129  
 AAM24129 standard; protein; 654 AA.

XX AAM24129;  
 AC  
 DT 17-OCT-2003 (revised)  
 DT 20-APR-1998 (first entry)  
 XX

DE *Pyrococcus furiosus* protease PFUS.

XX Protease; research reagent; thermal stability; *pyrococcus furiosus*.

XX *Pyrococcus furiosus*; DSM-3638.

XX WO9721823-A1.

XX 19-JUN-1997.

XX 07-NOV-1996; 96WO-UP003253.

XX 12-DEC-1995; 95UP-00323285.

XX (TAKI) TAKARA SHUZO CO LTD.

XX Takakura H, Moriehita M, Yamamoto K, Mita M, Asada K;

XX Tsunashima S, Kato I;

XX MPI; 1997-332794/30.

XX N-PSDB; AAT85695.

XX Protease(s) and gene encoding them obtained from *Thermococcus* and

XX *Pyrococcus* strains - have extremely high thermal stability and are useful

XX industrially and as research reagents.

XX Disclosure; Page 125-128; 159pp; Japanese.

XX This sequence represents the protease from *Pyrococcus furiosus* DSM-3638.

XX This sequence is a protease of the invention. The proteases can be

XX used as research reagents, and industrially in the food, drug and

XX chemical industries. (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 654 AA;

Query Match 72.3%; Score 2483.5; DB 2; Length 654;

Best Local Similarity 72.1%; Pred. No. 2.6e-144;  
 Matches 481; Conservative 69; Mismatches 96; Indels 21; Gaps 8;

```

QY 1 MKRLGAVLVALLVGLGALTAALPAVKPV--VRNNAVQKNGYGLTPGIFKVKQVMNNQ 58
DB 1 MKRLKALILVILVGLGVDSVVAAPKPKYQVRN--VEKNYGLLTPGLFKIKQKLNPNH 57
QY 59 EVDTVIMFGSYGDRDAVAVVLRMLGAQVYKXKIIPAVVVKIKARDLLIAGMIDTGYRG 118
DB 58 EISTVIVFENHREKEIAVRYLELGAQVRYVHIIIPALADLKVRLDLVLSGL--TG--G 113
QY 119 NTRVSGIKFIQEDYKYQVDDA-----TSVSOIGADTVNNSLGYDGSQVVAIVDTGIDAN 173
DB 114 KAKISGVRFIQEDYKYVSALELGLDESAQVMAITVMN-LGYDGSGITIGIIDTIDIAS 172
QY 174 HPDLKGVIGWYDAVNGRSTPYDDQGHGTHVAGVAGTSV-NSQYIGVAPGAKLVGVVVG 232
DB 173 HPDLQGVIGWYDFVNGRSTPYDDHGHGHVSAIAGTGAANSKKGKGAAPAKLVGIVV 232
QY 233 LGADSGSYSTIIAGVWVYVONKDKGIRVYNISLGSQSSQSDGTSLSOAVNNAMDAIIV 292
DB 233 LGADSGSSISTIIKGYEMAVDNNDKYGIRVYNISLGSQSSQSDGTSLSOAVNNAMDAIIV 292
QY 293 VVVAAGNSGPNYTYVGSPPAAASKVITVGAVDSDNINASFSSRGPADGR.LKEVVA PGVD 352
DB 293 VVVAAGNSGPNKYTVGSPPAAASKVITVGAVDKYDVTISFSSRGPADNRLKPEVVA PGW 352
QY 353 IIPRASGTSMGTPINDYTYTKASGTSMAIPIHVSQVALLIQAPHSWTPDKVKTAL.IETAD 412
DB 353 IIPRASGTSMGQPINDYTYTAAPGTSMAIPIHVSQVALLIQAPHSWTPDKVKTAL.IETAD 412
QY 413 IIPASGTSMGQPINDYTYTAAPGTSMAIPIHVSQVALLIQAPHSWTPDKVKTAL.IETAD 472
DB 413 IIPASGTSMGQPINDYTYTAAPGTSMAIPIHVSQVALLIQAPHSWTPDKVKTAL.IETAD 472
QY 473 WDTGSSPIDLYLYDPNGNEVDYSYTAAYGFEKGYNNPTAGTWTVKVSYKGAANYQVDV 532

```



```

Db      473 MNANSDLDLYLDYDNGQVDYSYTAAYGFEKGYGNPTDGTWTIKVSYSGSANYQVUV 532
Qy      533 VSDGSLSSQSGGNPNPNPNPTPTDQTGTFGSVNDVWDSTDTFTMNNSGATKTTGDL 592
Db      533 VSDGSLSSQ-----PSSSPSPQPEPTVDAKTFQSDHYHYDRSDTFTMTNNSGATKTTGDL 587
Qy      593 TPDTSYNDLDLYLDYDNGNLVDRSTSSNSYEHVEYANPAAGTWTPLVVAYSTYGMADYOL 652
Db      588 VEDTGYHDLDLVLDYDNPQKLVDRSSPNSYEHVEYILTPAGTMYFLVVAITYTGMAYYEL 647
Qy      653 KAVVYYG 659
Db      648 TAKVYYG 654

RESULT 6
AAM94841
ID      AAM94841 standard; protein; 654 AA.
XX
AC      AAM94841;
XX
DT      06-MAY-1999 (first entry)
XX
DE      Hyperthermostable protease.
XX
KM      Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
XX      additive; drug; washing agent; foodstuff; chemical synthesis.
XX
OS      Pyrococcus furiosus.
XX
PN      WO9856926-A1.
XX
PD      17-DEC-1998.
XX
PF      04-JUN-1998; 98WO-JP002465.
XX
PR      10-JUN-1997; 97JP-00151969.
XX
PA      (TAKI ) TAKARA SHUZO CO LTD.
XX
PI      Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
DR      WPI: 1999-080907/07.
XX
DR      N-PSDB; AAX05929.
XX
PT      Recombinant hyperthermostable protease from Pyrococcus furiosus - and
PT      gene encoding it, for large scale production of the protease for
PT      industrial use.
XX
PS      Disclosure; page 60-63; 82pp; Japanese.
XX
XX      The invention relates to a hyperthermostable protease derived from a
XX      thermophilic bacterium (especially Pyrococcus furiosus). The protease has
XX      working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
XX      (optimum 6-8), and retains more than 90% of its activity after 8 hours at
XX      95 deg.C. The invention also provides gene sequences encoding a
XX      polypeptide of formula SIG-A1a-Gly-Gly-Aan-Pro, where SIG is a signal
XX      peptide from subtilisin, and PRO is the above protease. Host cells
XX      (especially Bacillus strains) transformed with vectors comprising the
XX      genes are used for the recombinant production of the protease. The
XX      hyperthermostable protease which can be prepared in quantity suitable for
XX      industrial use, can be used as an additive for drugs, washing agents and
XX      foodstuffs and for chemical synthesis
SQ      Sequence 654 AA;

Query Match      72.3%; Score 2483.5; DB 2; Length 654;
Best Local Similarity 72.1%; Pred. No. 2.6e-144;
Matches 461; Conservative 69; Mismatches 96; Indels 21; Gaps 8,
1 MKRLGAVVIALVVLGLAGTALAAPVKPV--VNNNAVQOKNYGLLTPGLFKTVQRNNMNO 58
|||:::||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

```

Db      1 MKGLKALLVILVLGLVVGSAABAPEKKVEQVRN---VEKNYGLLTPGLFRKIQKLNPE 57
Qy      59 EVDVTIMFGSYGDRRAVAVKVLMLGAQVYSYKIIIPAVAVKIKARDLLLIAGMIDTGYG 118
Db      58 EISTIVIEENHREKELIARVLELMQAKRYVTHIIPALADLKARDLVISGL--TG--G 113
Qy      119 NTRVSGIKFIQEDYKVOYDVA-----TSVQIQADTWNNSLQDGSQVVAIVDTGIDAN 173
Db      114 KAKLSGVRFIQEDKYVTASAELEGDESAAQMAATYVWN--LGYDSSGITIGIIDTGIDAS 172
Qy      174 HPDLKGTIKYTDVANGSTPTPDQGHGTHVAGIYAGTGSV--NSQYIGVAPGAKLVGVAV 232
Db      173 HPDLQGVYGVWDFVNGRSYPYDDHGHGTHVASIAAGTGAASNGKYGKAPGAKLAGIYV 232
Qy      223 LGADSGSVSTIIAGVDWVYQKDKYGI RVINLSLSSQSSQSDGTSLQAVNMANDAGIV 292
Db      223 LGADSGSISTIIKGVEMVQDKDKYGIKVINLSLSSQSSQSDGTSLQAVNMANDAGLV 292
Qy      293 VCVAAAGNSGPNITYTVGSPPAAASKVITVGAVDNDNIASFSSRGPTADGRLKEBVVAPGVD 352
Db      293 VVVAAGNSGPNKYTTIGSPAAASKVITVGAVDKYDVTISFSSRGPTADGRLKEBVVAPGVM 352
Qy      353 IIAAPASGTSWGTPINDYYTTKASGTSMAIPHVSGVGAIILOHBSWTPDKVKTALILETD 412
Db      353 IIAAPASGTSWGPINDYYTTAAPTSMATPHVAGIALLLOHBSWTPDKVKTALILETD 412
Qy      413 IYAPKEIADIVAGARVAVVYKIKYDDYAKLTFQTSVNDVWDSTDTFTMNNSGATKTTGDL 472
Db      413 IYKPEIADIAGARVAVVYKAINVDYAKLFTYGVANKGSQTHQFVLSGSAFVATLY 472
Qy      473 WDTGSSDIDLXYLDYDNGNEVDYSYTAAYGFEKGYGNPTAGTWTIKVVSYSKGAANYQVDV 532
Db      473 MNANSDLDLYLDYDNGQVDYSYTAAYGFEKGYGNPTDGTWTIKVSYSGSANYQVUV 532
Qy      533 VSDGSLSSQSGGNPNPNPNPTPTDQTGTFGSVNDVWDSTDTFTMNNSGATKTTGDL 592
Db      533 VSDGSLSSQ-----PSSSPSPQPEPTVDAKTFQSDHYHYDRSDTFTMTNNSGATKTTGDL 587
Qy      593 TPDTSYNDLDLYLDYDNGNLVDRSTSSNSYEHVEYANPAAGTWTPLVVAYSTYGMADYOL 652
Db      588 VEDTGYHDLDLVLDYDNPQKLVDRSSPNSYEHVEYILTPAGTMYFLVVAITYTGMAYYEL 647
Qy      653 KAVVYYG 659
Db      648 TAKVYYG 654

RESULT 7
AAM24122
ID      AAM24122 standard; protein; 522 AA.
XX
XX      AAM24122;
XX
AC      AAM24122;
XX
DT      17-OCT-2003 (revised)
XX
DT      20-APR-1998 (first entry)
XX
DE      Pyrococcus furiosus protease.
XX
XX      Protease; research reagent; thermal stability; pyrococcus furiosus.
XX
OS      Pyrococcus furiosus; DSM-3638.
XX
XX
FH      Key
FT      Location/Qualifiers
FT      Misc-difference 428
FT      /label= Gly, Val
FT      /note= "encoded by GNA"
XX
XX
PN      MO9721823-A1.
XX
PD      19-JUN-1997.
XX
PF      07-NOV-1996; 96WO-JP003253.
XX

```

PR 12-DEC-1995; 95JP-00323285.  
 XX (TAKI ) TAKARA SHUZO CO LTD.  
 XX Takakura H, Morishita M, Yamamoto K, Mitra M, Aaeda K;  
 XX Teunasaawa S, Kato I;  
 XX WPI, 1997-332794/30.  
 DR N-PSDB; AAT85668.  
 XX  
 XX Protease(s) and genes encoding them obtained from Thermococcus and  
 PT Pyrococcus strains - have extremely high thermal stability and are useful  
 PT industrially and as research reagents.  
 XX  
 PS Claim 5; Page 87-90; 159pp; Japanese.  
 XX  
 XX This sequence represents the protease from Pyrococcus furiosus DSM-3638.  
 CC This sequence is a protease of the invention. The proteases of the  
 CC invention have extremely high thermal stability. The proteases can be  
 CC used as research reagents, and industrially in the food, drug and  
 CC chemical industries. (Updated on 17-OCT-2003 to standardise OS field)  
 CC  
 XX  
 SQ Sequence 522 AA;

Query Match 62.2%; Score 2138.5; DB 2; Length 522;  
 Best Local Similarity 78.3%; Pred. No. 3.3e-123;  
 Matches 407; Conservative 36; Mismatches 70; Indels 7; Gaps 3;

QY 141 SVSQIGADTVWNSLGYDGSQVVAIVDTGIDANHPDLKGVIGYDAVNGRSTPYDDQH 200  
 DB 9 SAAQVMATYVMN-LGYDGSQGITIGIIDTGDASHPDLOGKVIQWDFVNGSYPYDDHG 67  
 QY 201 GTHVAGIVAGTGSV-NSQYIGVAPGAKLVGKVLGADSGSVSTIIAGVDMVQNKXKG 259  
 DB 68 GTHVASTIAAGGAASNGKYGMAFGAKLAGIKVLGADSGSISTIIKGVMAVDNKKXKG 127  
 QY 260 IRVINTLSIGSSQSDGDTLSQAVNNAMDAGIIVCAAGNSGPTTYVSGSPAASKVITY 319  
 DB 128 IKVINTLSIGSSQSDGDTLSQAVNNAMDAGIIVCAAGNSGPTTYVSGSPAASKVITY 187  
 QY 320 GAVDSNDNIASFSSRGPTADGRLKPEVVAAGVDIIAPRASGTMGTPIINDYTTKASGTM 379  
 DB 188 GAVDKYDVTISFSSRGPTADGRLKPEVVAAGVDIIAPRASGTMGTPIINDYTTKASGTM 247  
 QY 380 ATPHVSQVGAIILOAHPSWTPDKYKTLIETADIVAPKETADIVAGRVVYAIKXDD 439  
 DB 248 ATPHVAAGIAALLLOAHPSWTPDKYKTLIETADIVAPKETADIVAGRVVYAIKXDD 307  
 QY 440 YAKLFTGSVADKGSATHTPDVSGATFVTATLYMDTSSDIDLTLYPDNGNEVDYSYTA 499  
 DB 308 YAKLVFTGVANKSQTHQFVIGASFTVATLYMDNANSDDLTLYPDNGNQVDYSYTA 367  
 QY 500 YGFERYGYNFTAGTWTYKVSYSKCAANYQVDVYSDGSLSSGGGNPNPNPNPTPTD 559  
 DB 368 YGFERYGYNFTAGTWTYKVSYSKCAANYQVDVYSDGSLSSGGGNPNPNPNPTPTD 422  
 QY 560 TQTFPGSNDVWDSTDTFMANVNSGATKITGDLFTSYNDLTLYPDNGNLVDRSTSS 619  
 DB 423 AKTFXSDHYHYDRSDITFTVNSGATKITGDLFTSYNDLTLYPDNGNLVDRSHSP 482  
 QY 620 NSYEHVEYANPAPGTWTELVVAASYTYGNADYQAKAVVYTG 659  
 DB 483 NSYEHVEYANPAPGTWTELVVAASYTYGNADYQAKAVVYTG 522

RESULT 8  
 AAM94838  
 ID AAM94838 standard; protein; 522 AA.  
 XX  
 AC AAM94838;  
 XX  
 XX 06-MAY-1999 (first entry)  
 DT  
 XX

DE Hyperthermostable protease.

XX Hyperthermostable; protease; thermophilic; bacterium; subtilisin;  
 KM additive; drug; washing agent; foodstuff; chemical synthesis.

XX Pyrococcus furiosus.

XX Key Location/Qualifiers  
 FT Misc-difference 428  
 FT /label= Gly or Val

XX WO9856926-A1.

XX 17-DEC-1998.

XX 04-JUN-1998; 98WO-JP002465.

XX 10-JUN-1997; 97JP-00151969.

XX (TAKI ) TAKARA SHUZO CO LTD.

XX Takakura H, Morishita M, Shimojo T, Aaeda K, Kato I;  
 XX WPI, 1999-080907/07.

XX Recombinant hyperthermostable protease from Pyrococcus furiosus - and  
 PT gene encoding it, for large scale production of the protease for  
 PT industrial use.

XX Claim 1; Page 39-41; 82pp; Japanese.

XX The invention relates to a hyperthermostable protease derived from a  
 CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has  
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10  
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at  
 CC 95 deg.C. The invention also provides gene sequences encoding a  
 CC polypeptide of formula SIG-Ala-Gly-Gly-Ala-PRO, where SIG is a signal  
 CC peptide from subtilisin, and PRO is the above protease. Host cells  
 CC (especially Bacillus strains) transformed with vectors comprising the  
 CC genes are used for the recombinant production of the protease. The  
 CC hyperthermostable protease which can be prepared in quantity suitable for  
 CC industrial use, can be used as an additive for drugs, washing agents and  
 CC foodstuffs and for chemical synthesis

XX Sequence 522 AA;

Query Match 62.2%; Score 2138.5; DB 2; Length 522;  
 Best Local Similarity 78.3%; Pred. No. 3.3e-123;  
 Matches 407; Conservative 36; Mismatches 70; Indels 7; Gaps 3;

QY 141 SVSQIGADTVWNSLGYDGSQVVAIVDTGIDANHPDLKGVIGYDAVNGRSTPYDDQH 200  
 DB 9 SAAQVMATYVMN-LGYDGSQGITIGIIDTGDASHPDLOGKVIQWDFVNGSYPYDDHG 67  
 QY 201 GTHVAGIVAGTGSV-NSQYIGVAPGAKLVGKVLGADSGSVSTIIAGVDMVQNKXKG 259  
 DB 68 GTHVASTIAAGGAASNGKYGMAFGAKLAGIKVLGADSGSISTIIKGVMAVDNKKXKG 127  
 QY 260 IRVINTLSIGSSQSDGDTLSQAVNNAMDAGIIVCAAGNSGPTTYVSGSPAASKVITY 319  
 DB 128 IKVINTLSIGSSQSDGDTLSQAVNNAMDAGIIVCAAGNSGPTTYVSGSPAASKVITY 187  
 QY 320 GAVDSNDNIASFSSRGPTADGRLKPEVVAAGVDIIAPRASGTMGTPIINDYTTKASGTM 379  
 DB 188 GAVDKYDVTISFSSRGPTADGRLKPEVVAAGVDIIAPRASGTMGTPIINDYTTKASGTM 247  
 QY 380 ATPHVSQVGAIILOAHPSWTPDKYKTLIETADIVAPKETADIVAGRVVYAIKXDD 439  
 DB 248 ATPHVAAGIAALLLOAHPSWTPDKYKTLIETADIVAPKETADIVAGRVVYAIKXDD 307  
 QY 440 YAKLFTGSVADKGSATHTPDVSGATFVTATLYMDTSSDIDLTLYPDNGNEVDYSYTA 499  
 DB 308 YAKLVFTGVANKSQTHQFVIGASFTVATLYMDNANSDDLTLYPDNGNQVDYSYTA 367

QY 500 YGEEKVGYNPTAGTWTWKVSYKGAANYQDVVSDGSLSSGGGNNPNPNPTPTTD 559  
 DB 368 YGEEKVGYNPTDGTWTWKVSYSGSANYQDVVSDGSLSSG-----RGSSPSPPPTPTVD 422  
 QY 560 TQFFTSQVNDYMDTSDTFPMVNSGATKIGTGLTFPTDSYNDLILYDPMGLVDRSTSS 619  
 DB 423 AKTFQXSDHYYSRSDTFPMVNSGATKIGTGLTFPTDSYNDLILYDPMGLVDRSESP 482  
 QY 620 NSEYHEVYANPAPGTWTFELVYASTYGMADYOLKAVYYG 659  
 DB 483 NSEYHEVYLTAPGTWTFELVYAYTYGMYELTAKVYIG 522

## RESULT 9

AAW94836  
 ID AAW94836 standard; protein; 412 AA.

AC AAW94836;

DT 06-MAY-1999 (first entry)

DE Hyperthermostable protease fragment.

KM Hyperthermostable; protease; thermophilic; bacterium; subtilisin;  
 additive; drug; washing agent; foodstuff; chemical synthesis.

OS Pyrococcus furiosus.

PN WO9856926-A1.

PD 17-DEC-1998.

PF 04-JUN-1998; 98WO-JP002465.

PR 10-JUN-1997; 97JP-00151969.

PA (TAKI ) TAKARA SHUZO CO LTD.

PI Takakura H, Morishita M, Shimajo T, Asada K, Kato I;

DR WPI; 1999-080907/07.

DR N-PADB; AAX05920.

PT Recombinant hyperthermostable protease from Pyrococcus furiosus - and  
 gene encoding it, for large scale production of the protease for  
 industrial use.

PS Claim 2; Page 35-37; 82pp; Japanese.

CC The invention relates to a hyperthermostable protease derived from a  
 thermophilic bacterium (especially Pyrococcus furiosus). The protease has  
 working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10  
 (optimum 6-8), and retains more than 90% of its activity after 8 hours at  
 95 deg.C. The invention also provides gene sequences encoding a  
 polypeptide of formula  $\text{StG-Ala-Gly-Gly-Asn-Pro}$ , where StG is a signal  
 peptide from subtilisin, and PRO is the above protease. Host cells  
 (especially Bacillus strains) transformed with vectors comprising the  
 genes are used for the recombinant production of the protease. The  
 hyperthermostable protease which can be prepared in quantity suitable for  
 industrial use, can be used as an additive for drugs, washing agents and  
 foodstuffs and for chemical synthesis

XX Sequence 412 AA;

QY Query Match 49.7%; Score 1707; DB 2; Length 412;  
 Best Local Similarity 80.6%; Pred. No. 8.7e-97;  
 Matches 325; Conservativity 29; Mismatches 47; Indels 2; Gaps 2;

QY 141 SVSGIGADTYWNSLIGYSGGVVAIVTGDANHPDLKGYIGYDAVNGRSTPYDQGH 200  
 DB 9 SAQGVWATYWN-LGYDGSIGITIGIDTGDASHPDQKGVIGVNDVNGRSYFYDDGH 67

QY 201 GTHVAGIVAGTGV-NSQYIGVAPGAKLVGVKVLGADGGSGSVSTIAGVDMVVOVKRG 259  
 DB 68 GTHVAGIAGTGAASNGKYMGPAGAKLAGIKVLGADGGSGSISTIIKGVEMAVDNDKYG 127  
 QY 260 IRYINISLSSGSSDSDTSLSSQAVNNMADAGIVCVAAAGSGPNYTVGSPPAASKVITY 319  
 DB 128 IKVINISLSSGSSDSDTSLSSQAVNNMADAGLVVVAAGSGPNKTYTGSPPAASKVITY 187  
 QY 320 GAVDSDNINIAFSRSPRTADGRKPEVVAQVVDIIAPRASGTSMTPIINDYTKASGTSM 379  
 DB 188 GAVDKYDVITSESSRPTADGRKPEVVAQVVDIIAPRASGTSMTPIINDYTKASGTSM 247  
 QY 380 ATPVSGVALLIQAPSWTPDKVKTALIEFDIAVAPKEIADIAYAGRVVYKAIKYDD 439  
 DB 248 ATPVAGIALLIQAPSWTPDKVKTALIEFDIAVAPKEIADIAYAGRVVYKAIKYDD 307  
 QY 440 YAKLTFTGSVADGSAITFTDVGATFVYATILYMDGSSSIDLYDPNENEDYSTAY 499  
 DB 308 YAKLTFTGVVANGSQTHQFVIGSASFVATILYMDANSDDLILYDPNENEDYSTAY 367  
 QY 500 YGEEKVGYNPTAGTWTWKVSYKGAANYQDVVSDGSLSSQG 542  
 DB 368 YGEEKVGYNPTDGTWTWKVSYSGSANYQDVVSDGSLSSQG 410

## RESULT 10

AAW87009  
 ID AAW87009 standard; protein; 237 AA.

AC AAW87009;

DT 10-MAY-1996 (first entry)

DE Hyperthermostable protease.

KM Protease; hyperthermostable; thermostability.

OS Pyrococcus furiosus.

PN WO9534645-A1.

PD 21-DEC-1995.

PF 05-JUN-1995; 95WO-JP001095.

PR 13-JUN-1994; 94JP-00130236.

PR 26-JUL-1994; 94JP-00173912.

PA (TAKI ) TAKARA SHUZO CO LTD.

PI Mita M, Yamamoto K, Morishita M, Asada K, Teunasa S, Kato I;

DR WPI; 1996-049674/05.

DR N-PADB; AAT08131.

PT Pyrococcus furiosus hyperthermostable protease gene - useful for  
 recombinant prodn. of hyperthermostable protease.

PS Example 3; Page 70-71; 85pp; Japanese.

CC The invention relates to the hyperthermostable protease of Pyrococcus  
 furiosus and its prodn. as a recombinant protein in transformants using a  
 vector carrying the protease gene (AAT08131). A genomic DNA sequence of  
 the invention is given in AAT08131 and its encoded protein in AAW87009

XX Sequence 237 AA;

QY Query Match 33.6%; Score 1154; DB 2; Length 237;  
 Best Local Similarity 97.4%; Pred. No. 4.7e-63;

QY 9 SAQGVWATYWN-LGYDGSIGITIGIDTGDASHPDQKGVIGVNDVNGRSYFYDDGH 67

Matches 224; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 176 DLKGVIGWYDAVNRSTPYDDQGHGTIVAGTGSVNSQYIGVAPGAKLVGVKVLGA 235  
 DB 1 DLKGVIGWYDAVNRSTPYDDQGHGTIVAGTGSVNSQYIGVAPGAKLVGVKVLGA 60

QY 236 DGSQSVSTIIAGVDMVYQNKDKYIRVYNLSIGSSQSDGTDLSQAVNNAMDAGIVVCV 295  
 DB 61 DGSQSVSTIIAGVDMVYQNKDKYIRVYNLSIGSSQSDGTDLSQAVNNAMDAGIVVCV 120

QY 296 AAGNGPMTYTVGSPAAASKVITVGAVDNDNIASFSSRGPTADRLKPEVVAPEGVDIIA 355  
 DB 121 AAGNGPMTYTVGSPAAASKVITVGAVDNDNIASFSSRGPTADRLKPEVVAPEGVDIIA 180

QY 356 PRASGTMGTPIINDYTTKASGTMATPRVSGVGLIIQAHPSMTPEDKYKT 405  
 DB 181 PRASGTMGTPIINDYTTKASGTMATPRVSGVGLIIQAHPSMTPEDKYKT 230

RESULT 11  
 AAR87011  
 ID AAR87011 standard; peptide; 188 AA.  
 XX  
 AC AAR87011;  
 XX  
 DT 10-MAY-1996 (first entry)  
 XX  
 DE Peptide sequence.  
 XX  
 KW Protease; hyperthermostable; thermostability.  
 XX  
 OS Pyrococcus furiosus.  
 XX  
 PN W09534645-A1.  
 XX  
 PD 21-DEC-1995.  
 XX  
 PE 05-JUN-1995; 95MO-JP001095.  
 XX  
 PR 13-JUN-1994; 94JP-00130236.  
 PR 26-JUL-1994; 94JP-00173912.  
 XX  
 PA (TAKI ) TAKARA SHUZO CO LTD.  
 XX  
 PI Mitra M, Yamamoto K, Morishita M, Asada K, Tsunawawa S, Kato I;  
 XX  
 DR WPI; 1996-049674/05.  
 XX  
 DR Pyrococcus furiosus hyperthermostable protease gene - useful for  
 PT recombinant prodn. of hyperthermostable protease.  
 XX  
 PS Example 4; Page 66-67; 85pp; Japanese.  
 XX  
 CC The invention relates to a gene (AAT08141) that codes for a  
 CC hyperthermostable protease (AAR87007) of Pyrococcus furiosus. 2 DNA  
 CC sequences (AAT08133-34) are provided encoding peptides (AAR87010-11)  
 XX  
 SQ Sequence 188 AA;

Query Match 27.2%; Score 935; DB 2; Length 188;  
 Best Local Similarity 98.4%; Pred. No. 1e-49;  
 Matches 184; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 200 HGTIVAGTGSVNSQYIGVAPGAKLVGVKVLGADSGSVSTIIAGVDMVYQNKDKYK 259  
 DB 2 HGTIVAGTGSVNSQYIGVAPGAKLVGVKVLGADSGSVSTIIAGVDMVYQNKDKYK 61

QY 260 IRVINISLIGSSQSDGTDLSQAVNNAMDAGIVVCVAAAGNSGPTTYTGSPPAAASKVITV 319  
 DB 62 IRVINISLIGSSQSDGTDLSQAVNNAMDAGIVVCVAAAGNSGPTTYTGSPPAAASKVITV 121

QY 320 GAVDSNDNIASFSSRGPTADRLKPEVVAPEGVDIIAPRASGTMGTPIINDYTTKASGTM 379  
 DB 182 GAVDSNDNIASFSSRGPTADRLKPEVVAPEGVDIIAPRASGTMGTPIINDYTTKASGTM 379

DB 122 GAVDSNDNIASFSSRGPTADRLKPEVVAPEGVDIIAPRASGTMGTPIINDYTTKASGTM 181  
 QY 380 ATPHYSG 386  
 DB 182 ATPHYTG 188

RESULT 12  
 AAM24126  
 ID AAM24126 standard; protein; 188 AA.  
 XX  
 AC AAM24126;  
 XX  
 DT 17-OCT-2003 (revised)  
 DT 20-APR-1998 (first entry)  
 XX  
 DE Thermococcus protease fragment.  
 XX  
 KW Protease; research reagent; thermal stability; thermococcus celer.  
 XX  
 OS Thermococcus celer; DSM-2476.  
 XX  
 PN W09721823-A1.  
 XX  
 PD 19-JUN-1997.  
 XX  
 PE 07-NOV-1996; 96MO-JP003253.  
 XX  
 PR 12-DEC-1995; 95JP-00323285.  
 XX  
 PA (TAKI ) TAKARA SHUZO CO LTD.  
 XX  
 PI Takakura H, Morishita M, Yamamoto K, Mitra M, Asada K;  
 PI Tsunawawa S, Kato I;  
 XX  
 DR WPI; 1997-332794/30.  
 DR N-PSDB; AAT85676.  
 XX  
 PT Protease(s) and genes encoding them obtained from Thermococcus and  
 PT Pyrococcus strains - have extremely high thermal stability and are useful  
 PT industrially and as research reagents.  
 XX  
 PS Disclosure; Page 110-112; 159pp; Japanese.  
 XX  
 CC This sequence represents a fragment of the protease from Thermococcus  
 CC celer DSM-2476 (see AAM24121 for full length sequence). This sequence is  
 CC a fragment of the protease of the invention. The proteases of the  
 CC invention have extremely high thermal stability. The proteases can be  
 CC used as research reagents, and industrially in the food, drug and  
 CC chemical industries. (Updated on 17-OCT-2003 to standardise OS field)  
 XX  
 SQ Sequence 188 AA;

Query Match 27.2%; Score 935; DB 2; Length 188;  
 Best Local Similarity 98.4%; Pred. No. 1e-49;  
 Matches 184; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 200 HGTIVAGTGSVNSQYIGVAPGAKLVGVKVLGADSGSVSTIIAGVDMVYQNKDKYK 259  
 DB 2 HGTIVAGTGSVNSQYIGVAPGAKLVGVKVLGADSGSVSTIIAGVDMVYQNKDKYK 61

QY 260 IRVINISLIGSSQSDGTDLSQAVNNAMDAGIVVCVAAAGNSGPTTYTGSPPAAASKVITV 319  
 DB 62 IRVINISLIGSSQSDGTDLSQAVNNAMDAGIVVCVAAAGNSGPTTYTGSPPAAASKVITV 121

QY 320 GAVDSNDNIASFSSRGPTADRLKPEVVAPEGVDIIAPRASGTMGTPIINDYTTKASGTM 379  
 DB 122 GAVDSNDNIASFSSRGPTADRLKPEVVAPEGVDIIAPRASGTMGTPIINDYTTKASGTM 181

QY 380 ATPHYSG 386  
 DB 182 ATPHYTG 188

|                       |   |
|-----------------------|---|
| RESULT 13             |   |
| AB09483               |   |
| ID                    | AB09483 standard; protein; 545 AA.  |
| XX                    |   |
| AC                    | AB09483;  |
| XX                    |   |
| DT                    | 11-SEP-2003 (revised)   |
| DT                    | 31-MAY-2002 (first entry)   |
| XX                    |   |
| DE                    | T. yonseit subtilisin-like serine protease.                               |
| XX                    |   |
| KW                    | Subtilisin-like serine protease.  |
| XX                    |   |
| OS                    | Thermoaerobacter yonseiensis.   |
| XX                    |   |
| PN                    | KR2000072141-A.   |
| XX                    |   |
| PD                    | 05-DEC-2000.  |
| XX                    |   |
| PF                    | 04-AUG-2000; 2000KR-00045411.   |
| XX                    |   |
| PR                    | 04-AUG-2000; 2000KR-00045411.   |
| XX                    |   |
| PA                    | (KIM Y S.   |
| XX                    |   |
| PI                    | Chang HJ, Kim DH, Byun YR, Kim YS;  |
| XX                    |   |
| DR                    | WPI; 2001-298092/31.  |
| XX                    |   |
| XX                    | N-PSDB; ABL54900.   |
| PT                    | New DNA sequence of thermophilic protein decomposition enzyme and protein |
| PT                    | derived therefrom.  |
| PS                    | Claim 3; Page 8; 15BP; Korean.  |
| XX                    |   |
| CC                    | This sequence represents the protein sequence of the Thermoaerobacter     |
| CC                    | yonsei subtilisin-like serine protease of the invention. (Updated on 11-  |
| CC                    | SEP-2003 to standardise OS field)   |
| XX                    |   |
| SO                    | Sequence 545 AA;  |
|                       |   |
| Query Match           | 26.7%; Score 918.5; DB 4; Length 545;                                     |
| Best Local Similarity | 43.4%; Pred. No. 4e-48;   |
| Matches               | 207; Conservative 73; Mismatches 150; Indels 47; Gaps 11.                 |
| QY                    | 35 VQKXNYGLITGELFKYQRRMNGVDVIMFGS-YGPRDRAVKYLRMGACVKTSTYKI 93             |
| DB                    | 1 MDSNKKNIFFDLBQRLNKPDSSEFPVILITFNKEVSPADIFTIAKNIGKFNIRKRYKI 60           |
| QY                    | 94 PAVAVKIKARLDLLLAGMIDTGYFGNTRVSGIKFTQEDYKYQVDDATSVSQIGADTVNNS 153       |
| DB                    | 61 PSIAANLKTSQINVL-----SKLEIVKQLEYDBPVATLDTATKMFGITKASD 109               |
| QY                    | 154 LGYDSSGVVAIVDTGIDANHPDLK-GKVIQWYDAVNGRSTPYDDQGHGTVAIGTG 212           |
| DB                    | 110 FGVGKNTITAIIDTGIDGNHYDLSGKTIIGMKDPIINKTTPPYDNGHGTVAISIAAGTG 169       |
| QY                    | 213 SVNSGYIVAPGAKLVGYKVTGADSGSVSTIIAGVDWYVQNKYKIGRVINLSIGSSQS 272         |
| DB                    | 170 AGNSLYKRVAPDALIVGIGKIVLDANGSGSMSTVTAIGDWA VONKDYGIKVINLSIGTSTS 229    |
| QY                    | 273 SDGTDTLSQAVNNAMADAGIVVCVAAAGNSGPNITYTGVSPAAASKVITVGAV---DSNNI 328     |
| DB                    | 230 SDGDTSTSLAVNRAVDSGIVVVAAGNSGPACTTIGSGGALEKAITVAAMADVGLGRNL 289        |
| QY                    | 329 ASFSSRGFTADGRLEPEVAVAPGVDTIAPRASGTSMTPIINDYYTKASGTSMTPHVSGVG 388      |
| DB                    | 290 ASFSSRGFTADGRIRPDIAAPGYNTITAKANS-----VNGYVT-YSGTSMATPRVAGTV 342       |
| QY                    | 389 ALLIOAHNSWPPDKTKTALIEPADIVAKRELADIVAGGRVVYDAIKYDVAKLTFTFGS 448        |
| DB                    | 343 ALMLAANLPLPNDKNIIMSTAKSWGPS-KNVYGRGRDLGVEAIRVAG--NFRGN 397            |

|               |  |     |
|---------------|--|-----|
| Oy            | 449 VAD-----KGSATHTFDVSGATF-VTATY---WPTGSSDIDLVLTPD                      | 487 |
| Dd            | 398 NIDVPMHYISGYLPGSRYSPTWTFFMAITNTSYPIAITLIIPDMANNYPDPFIYLXDP           | 454 |
| <br>RESULT 14 |  |     |
| ID            | AAM13667 standard; protein; 734 AA.                                      |     |
| AC            | AAM13667;  |     |
| DT            | 06-OCT-1997 (first entry)  |     |
| DE            | Streptomyces viridosporus dhpa gene product.                             |     |
| KM            | asymmetric hydrolase; dhpa; 4-substituted-1,4-dihdropyridine;            |     |
| KW            | derivative; Streptomyces viridosporus; ester; chiral; synthesis;         |     |
| KV            | cardiovascular; treatment; hypertension; ischaemic heart disease.        |     |
| OS            | Streptomyces viridosporus.   |     |
| FH            | Key Location/Qualifiers  |     |
| FT            | Peptide 205..724   |     |
| FT            | /note='fragment of dhpa protein; see AAM13666'                           |     |
| PN            | WO9705243-A1.  |     |
| PD            | 13-FEB-1997.   |     |
| PE            | 30-JUL-1996; 96WO-JP002147.  |     |
| PR            | 31-JUL-1995; 95JP-00212975.  |     |
| PR            | 29-FEB-1996; 96JP-00067478.  |     |
| PA            | (SAOC ) MERCIAN CORP.  |     |
| PI            | Arisawa A, Matsu Fuji M, Tsuruta T, Dobashi K, Nakashima T;              |     |
| PI            | Ishiki K, Yoshioka T;  |     |
| DR            | WI: 1997-145682/13.  |     |
| N-PSDB:       | AAT61454.  |     |
| PT            | Asymmetric hydrolase gene derived from Streptomyces viridosporus - acts  |     |
| PT            | on 4-substituted-1,4-di:hydro:pyridine derivatives to produce chiral     |     |
| PT            | derivatives useful for synthesis of cardiovascular drugs.                |     |
| XX            | Claim 3; Page 49-55; 78pp; Japanese.                                     |     |
| XX            | This sequence is an asymmetric hydrolase encoded by the Streptomyces     |     |
| CC            | viridosporus dhpa gene. The enzyme acts on 4-substituted-1,4-            |     |
| CC            | dihdropyridine derivatives. The enzyme allows the efficient conversion   |     |
| CC            | of 4-substituted-1,4-dihdropyridine esters to chiral partially           |     |
| CC            | hydrolysed derivatives, for use in the synthesis of cardiovascular drugs |     |
| CC            | suitable for the treatment of e.g. hypertension and ischaemic heart      |     |
| CC            | disease  |     |
| XX            | Sequence 734 AA:   |     |
| SQ            |  |     |
| Oy            | Query Match 19.3%; Score 665; DB 2; Length 734;                          |     |
| Oy            | Best Local Similarity 31.2%; Pred. No. 2.3e-32;                          |     |
| Oy            | Matches 216; Conservative 84; Mismatches 269; Indels 124; Gaps 26;       |     |
| Dd            | 22 LAAPVKVVNRNNAVOQKNYGLTGFLPKVQORMNMNEVDIVIMFGSYGDRBRAVKVRL 81          |     |
| Dd            | 97 LPAPAAALVNSGKDRLRFDITTELG--KAATRNSOKOGSLKVIV--GYGAARAKAEVR 151        |     |
| Oy            | 82 MGAOVKTSYKIIPAVAVKIKARDLLIAGMLDTGYFGATRVSGIKFIQEDRYKVVDATS 141        |     |
| Dd            | 152 EAGLEARTLTLSINADVARTPHED--ASELMDAVTNBGRTRASGIAPHWLDCVRRALPTS 208     |     |
| Oy            | 142 VSQIGADTVNSSLGYDSGGVVALVDVGIDANHPDLKSKYIGMTDAVNGRSFP--YDDQG 199      |     |

```

Db      209 VGOIGAPAPAM-SAGYDCKGVKXIALVDTGVDTSHHPDLKNGVTA---SKNFAPAAGDKRYG 264
QY      200 HGTNVAIGVAGTGSVN-SQYITGAPGAKLVGKVKTLGADSGSGVSTTITAGVDVWQNKDKY 256
Db      265 HGTNVAISTAGTGAGSQSGKRYGVAAPAAIILNKGVLTDGSGFGDSIIITAGGEMAAQC--- 320
QY      259 GIRVYNLSLGSOSGSDGDTSLSOAVNN-AMDAGIWVCVAGNNGENTYVGSPPAAASKVI 317
Db      321 GADVNNMSLGGMDTPR-TDPLBAADVKLBAKGVFLAIAAGNEBP---SISPGSADAL 377
QY      318 TVGAVDSNDUNIASSSKRP-TADGRKAEPEVAPGVDTIAPRASGTSWGTPIND--YYTK 373
Db      378 TVGAVDDKDLADSSSTGPRLGKAIKPEVTAPGVDTIAASAEQNDIQEBGEBPAGYMT 437
QY      374 ASGTSMA2PHVSGVGAALLQAHPSMTPPDKYKALJETADIAPKEADIAGVAGVNVYK 433
Db      438 ISGTSMA2PHVAGAAALLKQHPDMTSALKAL-TGSTKGKR-YTFBEGSSRIADK 494
QY      434 A1K-----YDDYAKLFTGSGVADKSGATH 457
Db      495 ALQQTIVADPVSVSGVQOMPHTDDEPVTKOLTYENLGTODVTLKLTSTATDPKGAAPA 554
QY      458 TFDVSGATFYATILYMDTGSSDIDLVLDPNENEDVASYAYGPEKGYVNPAGWTV 517
Db      555 GFPTLGATITTVPA---QGSASVDMADIRLGTVDGAYSALVVAATGGGQVVRITAAAVQR 611
QY      518 KVVSY-----KGAANYQVDVYSDGSISSGCGANPNPNPNPNPTTDTQTF-- 563
Db      612 EVESYDVTVAHIGRDKPPTTEHITDILIGVAGISGRGYG-----APATIDATILRP 662
QY      564 TGS-VNDYMDSTPTFM-----NVANSGATK1GDTLFTPT-SYNDLDTLYVDPNQNV 613
Db      663 KGYLVDSWIAKFGILKKGIDIMVLQPKLSYKOL-TITLDARTTKAADIIVDPK--- 717
QY      614 DRSTSNSYEHZVYANPAPGVTTFVLVAYSYTG 646
Db      718 -----AKPLSAT---TGYTYDTAG 733

```

PR 29-FEB-1996; 96JP-00067478.  
XX  
XX (SAOC ) MERCIAN CORP.  
PA  
XX Arisawa A, Matsufuji M, Tsuruta T, Dobaishi K, Nakashima T;  
XX Isehiki K, Yoshiooka T;  
PI  
XX WPI; 1997-145682/13.  
DR N-PSDB; AAT61455.  
XX  
XX Asymmetric hydrolyase gene derived from Streptomyces viridosporus - acts  
PT on 4-substituted-1,4-dihydropyridine derivatives to produce chiral  
PT derivatives useful for synthesis of cardiovascular drugs.  
PT  
PS  
PS Claim 5; Page 37-43; 78pp; Japanese.  
XX  
XX This sequence is a fusion protein comprising Streptomyces viridosporus  
CC dhpA gene product (an asymmetric hydrolyase) which acts on 4-substituted-  
CC 1,4-dihydropyridine derivatives, and melanin from S. antibioticus. The  
CC DhpA enzyme allows the efficient conversion of 4-substituted-1,4-  
CC dihydropyridine esters to chiral partially hydrolysed derivatives, for  
CC use in the synthesis of cardiovascular drugs suitable for the treatment  
CC of e.g. hypertension and ischemic heart disease. (Updated on 17-Oct-2003  
CC to standardise OS field)  
XX  
XX Sequence 823 AA;  
SQ

```

QY 22 LAAPKPVYRNAVNAQOKKNGVLLTPELKKYQKQRMNMOEVDVTVMFSYQDRDRAKYVLR 81
Db 97 LPADARLVASGKLEKRLRPDITELG--KAAATRSQOKGLVY--GYGAARPAKAEVR 151
QY 82 MGAQVYKSYKIIPAVAVYKIKARDULLAGMIDIGYFGNTRVSGIKFIQADYKVCYVDAT 141
Db 152 EAGELEKRLTTSINADAVRPHED---ASELMDAVTNGDRTASGIAHVWLDGVRBAALDTS 208
QY 142 VQOIGADVTVMNSLGYDGSQVVVAIVDPGIDIANHPDLKGYIKGYIDAVNRSTP--YDDG 199
Db 209 VQOIGAPPAAM--SAGIDGQGVVIAVLDPGVTSHPLDKGRYTA---SKNTAAFGADRYG 264
QY 200 HGTHTVAGIVAGTSVN--SQYIGVAPGAKLVGVVYLGADSGSVSTIIAGVDVWVQNKDY 258
Db 265 HGTHTVASIAAGTGAQSKKXYGVAPGAILNGKVLDDSGFGDSSGILAEWMAAAQ---- 320
QY 259 GIRVNLFLSSQSSSDGDTDSLQAVNN--AMDAGIVCVANAUNSGPNYTVSGPAAASVYI 317
Db 321 GADVNNMSLGMQDTPR--TDPLEAAVVDLSAEKVLPAIAAGNEGPB--SIGPGSADAL 377
QY 318 TVGAVDSNDNIASFSSRGP--TADGRLEKEVVAAGVPIIAPRASGSMGPIIND---YTTK 373
Db 378 TVGAVDDKDKLADPSSSTGPRIGDGAIPDVTAAQGVITAAASEGNDIQGEVGEAGIMT 437
QY 374 ASGTSMATPEHYSGVGAILLQAPHSWTPDKYKTLIETADIAPKEIADIAGAGRVNYK 433
Db 438 ISGTSMATPEHVAAGAAALLKQHPDMTSABLKGAL--TGSTKGAK--YTPPEQSSGRIQADK 494
QY 434 AIK-----YDGYAKLTFGGSVADGSAATH 457
Db 495 ALQGVVADPVSFVGVOQMPHHTDDEYTKQLTYRRLGQDVTALKLTIATDPRKKAAPA 554
QY 458 TFDVSGATFTATLYMDTSSDIDLVLDPNGNEVDYSTAYYGEKVGYNPAGTWTV 517
Db 555 GFTFLGATVTVPA---GSGASVDMTADRLRGTVDGAASAVVAVTGGQGTVYRTAAVQR 611
QY 518 KVSFY-----KGAANYOVADVVSQSSLSQSGCGAPNPNPAPNPTPTDQTF--- 563
Db 612 EYESIDVTYVRIHIGDKPTEHEHLDLIGVAGLSGRGY-----APATDTATLRLP 662
QY 564 TGS-VNDYMDTSDTFTM-----NVNSGATKTYTGDLTFDT-SYNDLIDLVLDPNGNLV 613

```

Db 663 KGTVLVDSWIANDFTLKGIDWLVOFKLSVTKDT-TLTLDARTTKADITVPPK----- 717  
Qy 614 DRSTSSNSYEHVEYANPAPGWTFTLVYASSTYG 646  
Db 718 -----AKPLSAT---IGTYVDTAG 733

Search completed: January 6, 2007, 22:24:53  
Job time : 144 secs

**This Page Blank (uspto)**



GenCore version 5.1.9  
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2007, 22:25:12 ; Search time 27 Seconds  
(without alignments)

2348.402 Million cell updates/sec

Title: US-10-800-684-1

Perfect score: 3437

Sequence: 1 MKRLGAVVLALVVLGLAGT.....YAVSTYGMADYQLKAVVYVG 659

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

PIR 80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                |
|------------|-------|-------------|--------|-------|----------------------------|
| 1          | 669.5 | 19.5        | 442    | 2     | A69587 intracellular alka  |
| 2          | 637   | 18.5        | 444    | 2     | B83891 intracellular alka  |
| 3          | 557.5 | 16.2        | 379    | 1     | SUBSCL subtilisin (EC 3.4  |
| 4          | 547   | 15.9        | 382    | 1     | subtilisin (EC 3.4         |
| 5          | 533.5 | 15.5        | 382    | 1     | SUBSN subtilisin (EC 3.4   |
| 6          | 533.5 | 15.5        | 1398   | 2     | T28159 pyrolysin (EC 3.4   |
| 7          | 524.5 | 15.3        | 274    | 1     | SUBSD subtilisin (EC 3.4   |
| 8          | 522.5 | 15.2        | 806    | 2     | A41341 microbisl serine p  |
| 9          | 519.5 | 15.1        | 275    | 2     | JC1085 subtilisin (EC 3.4  |
| 10         | 515.5 | 15.0        | 384    | 2     | JC4802 alkaline proteinase |
| 11         | 513.5 | 14.9        | 381    | 2     | JH0778 subtilisin (EC 3.4  |
| 12         | 512.5 | 14.9        | 381    | 1     | SUBSI subtilisin (EC 3.4   |
| 13         | 510.5 | 14.9        | 380    | 2     | A49778 high-alkaline seri  |
| 14         | 510   | 14.8        | 1433   | 1     | A36734 bacillopeptidase F  |
| 15         | 509.5 | 14.8        | 381    | 1     | SUBS8 subtilisin (EC 3.4   |
| 16         | 509.5 | 14.8        | 381    | 2     | JQ1487 subtilisin (EC 3.4  |
| 17         | 497.5 | 14.5        | 272    | 2     | A23624 subtilisin (EC 3.4  |
| 18         | 487   | 14.5        | 757    | 2     | C84120 subtilisin-type pr  |
| 19         | 495   | 14.4        | 374    | 2     | I39781 subtilisin (EC 3.4  |
| 20         | 491.5 | 14.3        | 378    | 2     | A33973 high-alkaline seri  |
| 21         | 488.5 | 14.2        | 279    | 1     | SUMYTV thermilase (EC 3.4  |
| 22         | 482   | 14.0        | 279    | 2     | I39974 serine proteinase   |
| 23         | 477.5 | 13.9        | 799    | 2     | G83753 subtilisin-type pr  |
| 24         | 473.5 | 13.8        | 535    | 2     | B82358 alkaline serine pr  |
| 25         | 473.5 | 13.8        | 715    | 2     | JC4508 alkaline serine pr  |
| 26         | 472   | 13.7        | 513    | 1     | A35742 equalysin (EC 3.4   |
| 27         | 470   | 13.7        | 361    | 2     | A48373 high-alkaline seri  |
| 28         | 464.5 | 13.5        | 534    | 1     | J50173 subtilisin-type al  |
| 29         |       |             |        |       | alkaline proteinase        |

|    |       |      |     |   |                            |
|----|-------|------|-----|---|----------------------------|
| 30 | 461   | 13.4 | 627 | 2 | D75393 serine proteinase,  |
| 31 | 452   | 13.2 | 440 | 2 | H72784 probable alkaline   |
| 32 | 445.5 | 13.0 | 519 | 2 | S71451 halolysin R4 (EC 3  |
| 33 | 443   | 12.9 | 488 | 2 | A11930 proteinase (import  |
| 34 | 437.5 | 12.7 | 372 | 2 | D83735 subtilisin-type al  |
| 35 | 437.5 | 12.7 | 645 | 1 | SUBSMP serine proteinase   |
| 36 | 432.5 | 12.6 | 397 | 2 | JM0075 cysteine-dependent  |
| 37 | 420.5 | 12.2 | 588 | 2 | C83836 subtilisin-type pr  |
| 38 | 418.5 | 12.1 | 401 | 2 | A57690 aerolysin precursor |
| 39 | 417   | 12.0 | 747 | 2 | T06580 subtilisin-like pr  |
| 40 | 412.5 | 12.0 | 467 | 1 | S45493 serine proteinase   |
| 41 | 407.5 | 11.9 | 745 | 2 | UC6119 subtilisin-like pr  |
| 42 | 399.5 | 11.6 | 326 | 1 | C41335 microbisl serine p  |
| 43 | 399   | 11.6 | 319 | 2 | I39866 alkaline serine p   |
| 44 | 398.5 | 11.6 | 525 | 2 | G84406 halolysin (importe  |
| 45 | 398   | 11.6 | 754 | 2 | T06579 subtilisin-like pr  |

#### ALIGNMENTS

##### RESULT 1

A69587 intracellular alkaline serine proteinase aprx - Bacillus subtilis

C/Species: Bacillus subtilis

C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #ext\_change 09-Jul-2004

C/Accession: A69587

R/Kmet, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertorello, C.; Bron, S.; Brouillet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chos

A.; Enllich, S.D.; Emerson, P.T.; Enllich, K.D.; Errington, C.; Ferrari, E. Nature 390, 249-256, 1997

A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;

Koetter, P.; Konings, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapins, A.; Lardinois, A.

A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maeda, S.; Maueel, Y.

M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, R.

Regier, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scallion, A.

A/Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,

akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Tetsura, P.; Tognoni, A.; Tosato, V.; Uchiyama, K

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K

A/Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A. A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A/Reference number: A69580; PMID:98044033; PMID:9384377

A/Accession: A69587

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-442 <KUN>

A/Cross-references: UNIPROT:Q11788; UNIPARC:UPI00006043F; GB:299113; GB:AL009126; NID:G

A/Experimental source: strain 168

C/Genetics:

A/Genes: aprX

F/146-398/Domain: subtilisin homology <SBT>

Query Match 19.5%; Score 669.5; DB 2; Length 442;  
Best local similarity 46.0%; Pred. No. 2.2e-29;

Matches 151; Conservative 51; Mismatches 103; Indels 23; Gaps 9;

|    |     |   |     |
|----|-----|---|-----|
| QY | 127 | FIQEDYKQVDDATVSQIGADTWNISLGYDGGVVAIVDTGIDANHPDLKQVIGWD    | 186 |
| DB | 116 | YINREVKALDPTATEASH-AKEVVRNGQTLTGKGVAVAVDGI-YPHDLEIRIIGFAD | 173 |
| QY | 187 | ANNGRSTPYDDGCHTHVAGIVAGGSVNS-QYIGVAPAKVGVVGLADSGSGSTII    | 245 |
| DB | 174 | MVNOQTEPYDDGCHTHVAGIVAGGSVNS-QYIGVAPAKVGVVGLADSGSGSTII    | 245 |
| QY | 246 | AGVDWVQ-NKDK--YGIKVINLSGSS--OSSDGTSLDQAIVNNAADAGIVVCVAA   | 299 |
| DB | 234 | EGVEMCIQYNEONPDEPIDIMSGDLARRDHEEDLVAVBEAMAGIVVCVAA        | 293 |
| QY | 300 | SGBNNTVGSPPAASRVITVGAVDN-----DNIASSSGKPTADGLKBEVAVPG      | 352 |
| DB | 294 | SGDSDQTIASPGVSEKVTIVGALDDNMTASDDDTVAFSSRSGPTVYGKEXDIIAP   | 353 |



C:Keywords: hydrolase; serine proteinase  
F:136-342/Domain: subtilisin homology <SBT>

Query Match 15.9%; Score 547; DB 2; Length 382;  
Best Local Similarity 34.5%; Pred. No. 8.8e-23;

Matches 154; Conservative 73; Mismatches 145; Indels 74; Gaps 17;

QY 1 MKRLGAVLV-ALVVLGLAGTALAPKPVVNNAAVQKXVGLTTPGLFKVQRMNN 57  
DB 1 MKKLFTKVAASALLSLTATSVSAE-----EQKQYLI--GFENQLQVTERV 48  
QY 58 QEVN---TVIMFGSYGRDRAVKVRLMLGAQVKSXKIIIPAVAKIKARDLLIAGMID 113  
DB 49 EESDKQSSMSLFAEVNDESIEMELL-----YEFEDIPVVSVELSPEDVDL----- 95  
QY 114 TGYFGNTRVSGIKFIQEDYKNOVDAT---VSQIGADTVNMSLGDGSGVVAIVDTGI 170  
DB 96 -----EKQPSITYIEDIEVITITNQVTWGTTRVQAPFAM--TRGYTGCVRAVADTGI 148  
QY 171 DANHPDLKGVIGWYDAVNGRSTPYVDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGV 230  
DB 149 -STHPDL--NIRGCVSFPVGEPSYDQNGHGHVAGTIALNN-SIGVGVAPNAELIYAV 204  
QY 231 KYLGADSGSVSTIIAGVDMVYQNKXGIRVNLISGSSQSDGTDLSQAVNNAMADG 290  
DB 205 KYLGANGSGSVSIAQGLQMTAQN-----NIHVANISLGSPPV--GSQTELVAVNQATNAG 257  
QY 291 IYVCAVANGSGPNTTVSGPAAASKVITYGAVDSNDNIASFSSRGPTADGRLEKPEVAVG 350  
DB 258 VLVVAVATGNGSG--TVSYPARYANALAVGATDQNNNRASFQOYGTGLN-----IVAGG 309  
QY 351 VIIIPRAGTSMGPINNYTKASGTMATPHVSGVGLIIQAHPSWTPDKYKALIEFT 410  
DB 310 VCI-----QSTYPRKRYASLSGTSMTATPHVAGVAAALVKQKPSMNTQIRQLIST 360  
QY 411 ADIVAPKEIADIAYAGRVNYYKAIK 436  
DB 361 ATSLGNSN---QFGSLVNAEAATR 382

RESULT 5  
SUBSN  
subtilisin (EC 3.4.21.62) BPN' precursor - Bacillus amyloliquefaciens  
N:Alternate names: subtilisin Novo  
C:Species: Bacillus amyloliquefaciens  
C:date: 24-Apr-1984 #sequence\_revision 28-Aug-1985 #text\_change 05-Oct-2004  
C:Accession: B25415; A93495; T44584; A92033; A00970  
R:Vaanantha, N.; Thompson, L.D.; Rhodes, C.; Banner, C.; Nagle, J.; Filipula, D.  
J. Bacteriol. 159, 811-819, 1984  
A:Title: Genes for alkaline protease and neutral protease from Bacillus amyloliquefacient  
A:Reference number: A25415; MUID:85006739; PMID:6090391  
A:Accession: B25415  
A:Molecule type: DNA  
A:Residues: 1-382 <VAS>  
A:Cross-references: UNIPROT:P00782; UNIPARC:UPI0000136180; GB:K02496; NID:g142525; PIDN:  
A:Experimental source: ATCC 23844  
R:Wells, J.A.; Ferrari, E.; Henner, D.J.; Estell, D.A.; Chen, E.Y.  
Nucleic Acids Res. 11, 7911-7925, 1983  
A:Title: Cloning, sequencing, and secretion of Bacillus amyloliquefaciens subtilisin in  
A:Reference number: A93495; MUID:84069812; PMID:6316278  
A:Accession: A93495  
A:Molecule type: DNA  
A:Residues: 1-382 <WEL>  
A:Cross-references: UNIPARC:UPI0000136180  
A:Accession: T44584  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 'M', 8-382 <ME2>  
A:Cross-references: UNIPARC:UPI0000136180; EMBL:X00165; NID:g39337; PIDN:CAA24990.1; PID  
R:Markland, F.S.; Smith, E.L.  
J. Biol. Chem. 242, 5198-5211, 1967  
A:Title: Subtilisin BPN'. VII. Isolation of cyanogen bromide peptides and the complete a  
A:Reference number: A92033; MUID:68086682; PMID:6065094

A:Accession: A92033  
A:Molecule type: protein  
A:Residues: 108-162, 'PN', 165-167, 'D', 169-194, 'SA', 197-204, 'DA', 207-264, 'ST', 267-357, 'Q',  
A:Cross-references: UNIPARC:UPI000011292B  
R:Kratz, J.  
In The Enzymes, 3rd ed., vol.3, Boyer, P.D., ed., pp.547-560, Academic Press, New York, 1

A:Title: Subtilisin: X-ray structure.  
A:Reference number: A94443  
A:Contents: annotation; X-ray crystallography, 2.5 angstroms; active site  
A:Comment: Secretion of subtilisin is associated with the onset of sporulation, and many  
not necessary for normal sporulation.

C:Genetics:  
A:Start codon: GTG  
C:Superfamily: Subtilisin; subtilisin homology  
C:Keywords: hydrolase; serine proteinase  
F:1-32/Domain: signal sequence #status predicted <SIG>  
F:33-107/Domain: activation peptide #status predicted <APT>  
F:108-382/Product: subtilisin BPN' #status experimental <MPT>  
F:130-342/Domain: subtilisin homology <SBT>  
F:139,171,328/Active site: Asp, His, Ser #status experimental

Query Match 15.5%; Score 533.5; DB 1; Length 382;  
Best Local Similarity 38.5%; Pred. No. 4.8e-22;  
Matches 143; Conservative 52; Mismatches 113; Indels 63; Gaps 14;

QY 78 VRLMGAQVKSXKIIIPAV-----AVKIKARDLLIAGMIDTGYFGNTRVSGIKFIQ 130  
DB 59 VISEKGVQKQKPYDADAATLNEKAYELKD-----PSVAYVE 100  
QY 131 DYKQVQVDAT---SVSQIGADTVNMSLGYDGSVVAIVDTGIDANHPDLKRGVIGWYDA 187  
DB 101 DHVAAVAVQSVYGVYQIKAPAL--HSQYGTGSNVVAVAVDSDIDSHPL--KVAGASAM 157  
QY 188 VNGRSTPY--DDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGVYKVLGADSGSVSTIIA 246  
DB 158 VSEETNPFQDNNSHGHVAGVAAALNN--SIGVLGVAPASLAVVYKVLGADSGSGQYSWIIIN 216  
QY 247 GVDWVYQNKXGIRVNLISGSSQSDGTDLSQAVNNAMADGIVVCAAGNSGP--NT 304  
DB 217 GTEMAIAN---NMVIVNNISLG--GPGSALAKVAADVAVSGVVAAGNBTGSSS 269  
QY 305 YTVSGPAAASKVITYGAVDSNDNIASFSSRGPTADGRLEKPEVAVGVIIIPRAGTSMG 364  
DB 270 STVGYPRKXPSVIANGADSSNQGRASFSSVGEELD-----VMAQVSI-----QS 314  
QY 365 TPINDYTTKASGTMATPHVSGVGLIIQAHPSWTPDKYKALIEFTADIAPKEIAD-IA 423  
DB 315 TLPGKRYGAVNGTSMASPHVAGAAALILSKHNMVTQVRSLENTT-----TKLGDSPFY 369  
QY 424 YGAGRVNYYKA 434  
DB 370 YGKGLINVOAA 380

RESULT 6  
T28159  
pyrolysin (EC 3.4.-.-) - Pyrococcus furiosus  
C:Species: Pyrococcus furiosus  
C:date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T28159  
R:Vochohorst, W.G.B.; Eggen, R.I.L.; Geerling, A.C.M.; Platceeuw, C.; Sieren, R.J.; Vos, Y  
J. Biol. Chem. 271, 20426-20431, 1996  
A:Title: Isolation and characterization of the hyperthermostable serine protease, pyrolyse  
A:Reference number: Z20481; MUID:96355370; PMID:8702780  
A:Accession: T28159  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1398 <VOO>  
A:Cross-references: UNIPROT:P72186; UNIPARC:UPI000016FD59; EMBL:U55835; NID:g1556462; PII  
A:Experimental source: DSM3638  
C:Genetics:  
A:Gene: pls  
C:Keywords: hydrolase; serine proteinase

Query Match 15.5%; Score 533.5; DB 2; Length 1398;  
 Best Local Similarity 22.6%; Pred. No. 2.7e-21;  
 Matches 239; Conservative 104; Mismatches 248; Indels 467; Gaps 37;

```

QY 2 KRIGAVVLAALVYGLAGTALAAPKPPVRRNNAVQ--KATGLTPTGLFKKYQQRNNN-- 57
DB 4 KGLVTFEITMLLSVPHVPHVPSAGTPPVSSSENSTTSLIPNOQVLT---KEVSQAALNAI 59
QY 58 ---QEVTVIMEGSGYDRBRAVKVLRLMGAQVKYSYKIIIPAVAKI---KARDLLIAGM 111
DB 60 MKGQNNMVLIIKTKGKLEAKTELEKAGIILDENRVLNMLVKIKPEKKELENYISL 119
QY 112 IDTGFGTRVSGIKFIEDYKVOVDATSVSQTGADTVNSL-----GYDSSGVVVAI 165
DB 120 EKAMINREVKLSP-PIVEKDVYTK-EPSELBPMTNMTWVINALQFIOEFGYDGGVVAV 177
QY 166 VDTGIDANH-----PDLKGVICMYD-----ANNG----- 190
DB 178 LDTGVDPNHPFLSTTPDGRKKIEMKQFTDGFVDTSPSSKVVNGTLLINTTFQVAGL 237
QY 191 ----- 190
DB 238 TINESTGMEVYVTVVYVNTIGNITSANGIHFGLLPERKYPDLNPDQEDPYVLLV 237
QY 191 RST----- 193
DB 298 NSTGNGVDIAVDTLDYDFTDEVLQGVNTVDVAVSYGYPGLNYLAIDPNGEYAV 357
QY 194 -PYDQGHGTHVAGIVAGTGSVN-----SOYI-----GVAP 223
DB 358 FGMGHGHGTHVAGTAVGYSNNDAWMLMSYSEWVFRLYGMDYNTVTTDVOGVAP 417
QY 224 GAKLVGVKVLGADSGSVSTIIAGVDWVONKRYGIRVNLSTL-SGSSQSSDGTDSLQX 282
DB 418 GAQIMARVRSIDRSGSMMDIIEGMTYAA---THGADVISMISLGNAAPVLDGTDEBSVA 473
QY 283 VNNAWMDA-GIVVCYCAAGNSGPNNTYTGSSPAAASKVITVGVAD----- 323
DB 474 VDELTETKYGAVFIAMNBGPINIVGSPVATKATITVGAAPVINGVYVSGALGYPPY 533
QY 324 -----SNDNIASFSSRGPTADGRLKEDEVAPGVDIAPRAGTSMQPTINDYTKA 374
DB 534 YGFYPPAYTNVRIAPFSSRGPRIDGRIKNVVAAPGVYI---SLPMWIGADP--M 586
QY 375 SGTSMAIPHYSGVIGALILQI-HPS--WTPDKVKYALIEA---DIVAPKEIADIAVG 425
DB 587 SGTSMAIPHYSGVIGALILQI-HPS--WTPDKVKYALIEA---DIVAPKEIADIAVG 425
QY 426 AGRVNY-----YKAI-----KYDDYAKL----- 443
DB 647 HGLVNTTKSMELLKALNGTTLPIVDHMAKSYSDFABYIGVDVIRGLVARNSPIDIVEMH 706
QY 444 -----TF-----TGSV----- 449
DB 707 IKTVGDTETRTFEIYATEPFIKPFVSGVILNNTTEFVLRVKVDGLEGLEGLYGRITID 766
QY 450 -----ADKGSATIT-EPVSGATITV-----ATLY 472
DB 767 DPTTVEIIEELIINTIVIPKFTPENNYTLTWYDINPEVTHHFTFVPGVDVLYAMTTY 826
QY 473 WDTSGLDILLYLDYDPNG-----NEVDYSTAYYGFPEKGVYNPPTAGTVVKKVYSKGAAN 527
DB 827 WDTG-----LYRPMGMFVFPYQDLYLAA-----VSNPMGNNMLVWTGTFNAPL 871
QY 528 YQVDVVSQSLQSGGNGNPNPNPNPTPTDQTGFTGSVN-DYWDTSPTMTNVA--- 582
DB 872 YR-----SGFLVRIRHGVETIPS-----VMYINRTYLDTNTFSEIENINNI 912
QY 583 -----SGATKITGDLTF-----DTSYNDLDYTL 605
DB 913 YAPINATLIPIGLGTYNASVSVSGDFEPIKIEVPEGTAEIKIRIGNPSVPSNLDLYL 972

```

QY 606 YDPNGNLVDRSTSSNSYEHVEYANPAGTWTFLVAYS 643  
 DB 973 YDSKGNLVALDGNPRTABEEVVEYPRGYSTIVHGYS 1010

## RESULT 7

## SUBSD

subtilisin (EC 3.4.21.62) DY - Bacillus subtilis (strain DY)  
 N/Alternate names: alkaline serine proteinase  
 C/Species: Bacillus subtilis  
 A/Variety: strain DY

C/Date: 20-Sep-1984 #sequence\_revision 20-Sep-1984 #text\_change 05-Oct-2004  
 C/Accession: A00969; S02492

R/Nedkov, P.; Oberthur, W.; Braunitzer, G.  
 Biol. Chem. Hoppe-Seyler 366, 421-430, 1985

A/Title: Determination of the complete amino-acid sequence of subtilisin DY and its comp  
 A/Reference number: A00969; PMID:85279896; PMID:3927935

A/Accession: A00969

A/Molecule type: protein

A/Residues: 1-274 <MED>

A/Cross-references: UNIPARC:UPI000002D020

R/Illova, A.; Kleinachmidt, T.; Nedkov, P.  
 Biol. Chem. Hoppe-Seyler 368, 1479-1487, 1987

A/Title: Reductive alkylation of lysine residues in subtilisin DY.  
 A/Reference number: S02492; PMID:88134577; PMID:3124865

A/Accession: S02492

A/Molecule type: protein

A/Residues: 1-212-177-22-24-27-29-43-45-93-95-123-135-137-140-142-169-171-183-187-221-236  
 A/Cross-references: UNIPARC:UPI0000055F8; UNIPARC:UPI0000055F8; UNIPARC:UPI00001638C;  
 BFI; UNIPARC:UPI0000172BF; UNIPARC:UPI0000172BF; UNIPARC:UPI0000172BF; UNIPARC:UPI0000172BF;  
 C/Comment: Secretion of subtilisin is associated with the onset of sporulation, and many  
 not necessary for normal sporulation.

C/Superfamily: Subtilisin; subtilisin homology  
 F/Keywords: extracellular protein; hydrolase; serine proteinase  
 F/32-33/Domain: subtilisin homology <ST>

F/32,63,120/Active site: Asp, His, Ser #status predicted

Query Match 15.3%; Score 524.5; DB 1; Length 274;  
 Best Local Similarity 42.2%; Pred. No. 9.5e-22;  
 Matches 132; Conservative 38; Mismatches 94; Indels 49; Gaps 10;

QY 124 GIKFTQEDYKVOVDATSVSQTGADTVNSLGYDSSGVVAIVDTGIDANHPDLKGVYIG 183  
 DB 7 GIPLIKAD-KYQAAQ-----GYKANVKGIIPTGIAAHTDL--KVVG 46

QY 184 WYDAYNGRSTYDDQGHGTHVAGIVAGTGSVNSQYIVGAPAKLVGVVYLGADSGSYST 243  
 DB 47 GASFTVSGESTYNTDNGHGHVAGTVAAALDNTTG-VLGAAPVVSLYAIKVLNSGSGTYSA 105

QY 244 IYAGVWVONKDKKXGIRVINLSGSSQSSDGTDSLQAVNNAMDAIVVCVAAAGNSGP- 302  
 DB 106 IVSGIEMATON-----GLDIVIMSLG---GPGSGTALQADVKAISGIVVVAALAGNSGS 158

QY 303 -NTTVGSPAAASKVITYIGAVDSNDNIASFSSRGPTADGRLKEDEVAPGVDIAPRAGT 361  
 DB 159 GSQNTTIGYPAKYDVIIVGAVDSNKNRASFSSVG-----AELEVMAPGVSVSTYPSNT 212

QY 362 SMGTPIINDYKASGTATPVASGVGALLIQAPSPVTPDKVTKALLETADIAPKXIAD 421  
 DB 213 -----YTSINGTSMASHVAGAAALILSKYPTLSASQVNRRLSSTYATNIGD---S 259

QY 422 IYAGRVNYVYKA 434  
 DB 260 FYYGKGLINVEAA 272

## RESULT 8

A41341 serine proteinase (EC 3.4.21.-), minor (vpr), precursor - Bacillus subtilis  
 macrobia1 subtilis  
 C/Species: Bacillus subtilis  
 C/Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 09-Jul-2004  
 C/Accession: A41341; B41341; S59700; D69730



QY 303 -NTYVGSPPAASKVITVGAVDNDNIASFSSRGPTADGRLKEPVAVDIIAPRASGT 361  
 Db 160 GNTNITIGPAKVDSDIYAIGADNSNRASFSSVG-----ABELEMAG-----AGV 205  
 QY 362 SMGTPINDYTYKAGTSMATPHVSGVALIIQAPSWTPDKTALLETADIAPKEIAD 421  
 Db 206 YSTYETNTYAT-LNGTSMASEHVAAGAAALILSKHPNLSAQVNRRLSTATYLG-----SS 260  
 QY 422 IAYGAGRVVYKA 434  
 Db 261 FYYGKGLINVERA 273

## RESULT 10

JC4802  
 alkaline proteinase (EC 3.4.21.-) precursor - Thermactinomyces sp. (strain E79)  
 C/Species: Thermactinomyces sp.  
 A/Variety: strain E79  
 C/Date: 15-Aug-1996 #sequence\_revision 15-Oct-1996 #text\_change 05-Oct-2004  
 C/Accession: JC4802  
 R/Lee, J.K.; Kim, Y.O.; Kim, H.K.; Park, Y.S.; Oh, T.K.  
 Bioosci. Biotechnol. Biochem. 60, 840-846, 1996  
 A/Title: Purification and characterization of a thermostable alkaline protease from *Thermactinomyces*  
 A/Reference number: JC4802; MUID:96261070; PMID:8704314  
 A/Accession: JC4802  
 A/Molecule type: DNA  
 A/Residues: 1-384 <LBS>  
 A/Cross-references: UNIPROT:O56365; UNIPARC:UPI00000B2BBF; GB:U11759; NID:G1389689; PIND:  
 A/Experimental source: strain E79  
 C/Comment: This protein is thermostable.  
 C/Function:  
 A/Description: extracellular alkaline serine proteinase [validated, MUID:96261070]  
 C/Superfamily: Subtilisin; subtilisin homology  
 C/Keywords: hydrolase; serine proteinase  
 F/1-25/Domain: signal sequence #status predicted <SIG>  
 F/26-106/Domain: propeptide #status predicted <PRO>  
 F/107-383/Product: alkaline proteinase #status experimental <MAT>  
 F/134-344/Domain: subtilisin homology <SBT>  
 F/133-176/330/Active site: Asp, His, Ser #status predicted

Query Match 15.0%; Score 515.5; DB 2; Length 384;  
 Best Local Similarity 32.9%; Pred. No. 4.6e-21;  
 Matches 147; Conservative 77; Mismatches 150; Indels 73; Gaps 18;

QY 1 MKRLGAVTALVIVGLIA--GTALAAPKPEVFNNAVVOQKNGILTPGLFKVQVQANMNG 58  
 Db 1 MKRFLSVATALLLVLLAVPGTWFAA--SPASTDDVYPGLIVKFDGISAQ----- 50  
 QY 59 EVDTVIMFGSGYGRDRAVKRLRLMGAQVYKSYKIIPAVAVKIKARDLLIAGMIDTGYFG 118  
 Db 51 --STGSIHAQYG--AKSIEKSKYLGEFV-----VKFDGS-----VEKKIEK-YKN 90  
 QY 119 NTFVSGIKFIQEDYKQV---DQATS---VSOIGADVTYNSLGYDGSQVVAIVDTGI 170  
 Db 91 NPVV---EYVERPHYHIMMTPRDLTSROMGPKQVAPQAMD-VTSSSSYVIAIVDTGV 146  
 QY 171 DANHPDLKGVIGVAVNAGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGALVGV 230  
 Db 147 QTNHPDLQGIYQGYFVNDNSNPQDNGHGTGACAGIAAATVNNGTGIAGMAPNMSIMPV 206  
 QY 231 KVLGAGSGSVSTIINGVDMVYVONKQKIGRVINILSGSSQSDGTDLSQAVNNAMDAG 290  
 Db 207 KVLNNGSGTMAVANGIYAAQN---GADVISLSLG---GTSGSALQSAVQQAAMNSG 259  
 QY 291 IYVCAVAGNSGPTTYVGSPPAASKVITVGAVDNDNIASFSSRGPTADGRLKEPVAVG 350  
 Db 260 AYVVAAGNS--SSSTPNYPAYYSQALIAVASTDSNLSYFSNYGKWD-----VAAPG 311  
 QY 351 VVLIAPRASGTSMGTPINDYTYKAGTSMATPHVSGVALIIQAPSWTPDKTALLET 410  
 Db 312 SNIYS-----TYLNSYASLSGTSMATPHVAGIALL--ASQGRNSQIRAIAMENT 360

QY 411 ADVAPKEIADIVGAGRVVYAIKY 437  
 Db 361 ADKISG---TGTVFQHRINAVYAVNY 384

## RESULT 11

JH0778  
 subtilisin (EC 3.4.21.62) NAT precursor - *Bacillus subtilis* (strain natto NC2-1)  
 N/Alternate names: natto proteinase; nattokinase; subtilisin BSP  
 C/Species: *Bacillus subtilis*  
 C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 05-Oct-2004  
 C/Accession: JH0778; J50601; J50517; JC2036  
 R/Nakamura, T.; Yamagata, Y.; Ichishima, E.  
 Bioosci. Biotechnol. Biochem. 56, 1863-1871, 1992  
 A/Title: Nucleotide sequence of the subtilisin NAT gene, aprN, of *Bacillus subtilis* (natto)  
 A/Reference number: JH0778; MUID:93113095; PMID:1369081  
 A/Accession: JH0778  
 A/Molecule type: DNA  
 A/Residues: 1-381 <NAK>  
 A/Cross-references: UNIPROT:P35835; UNIPARC:UPI000000AAB8; GB:D25319; NID:G435439; PIND:  
 R/Sumi, H.; Seibutsu 29, 119-123, 1991  
 Kagaku To Seibutsu 29, 119-123, 1991  
 A/Title: Natto kinase and fibrinolysins.  
 A/Reference number: J50601  
 A/Accession: J50601  
 A/Molecule type: protein  
 A/Residues: 107-381 <SUM>  
 A/Cross-references: UNIPARC:UPI00001565DB  
 R/Sumi, H.; Nakajima, N.  
 Nippon Nogei Kagaku Kaishi 65, 1125-1127, 1991  
 A/Title: Studies on fibrinolytic enzymes in fermentation food.  
 A/Reference number: J50517  
 A/Accession: J50517  
 A/Molecule type: protein  
 A/Residues: 107-381 <SU2>  
 A/Cross-references: UNIPARC:UPI00001565DB  
 R/Fujita, M.; Nomura, K.; Hong, K.; Ito, Y.; Asada, A.; Nishimuro, S.  
 Biochem. Biophys. Res. Commun. 197, 1340-1347, 1993  
 A/Title: Purification and characterization of a strong fibrinolytic enzyme (nattokinase)  
 A/Reference number: JC2036; MUID:94107337; PMID:8280151  
 A/Accession: JC2036  
 A/Molecule type: protein  
 A/Residues: 107-381 <FU>  
 A/Cross-references: UNIPARC:UPI00001565DB  
 C/Genetics:  
 A/Gene: aprN  
 A/Start codon: GTG  
 C/Superfamily: Subtilisin; subtilisin homology  
 C/Keywords: hydrolase; serine proteinase; zymogen  
 F/1-23/Domain: signal sequence #status predicted <SIG>  
 F/30-106/Domain: activation peptide #status predicted <PRO>  
 F/107-381/Product: subtilisin NAT #status experimental <MAT>  
 F/129-341/Domain: subtilisin homology <SBT>  
 F/138,139,170,327/Active site: Asp, Ser, His, Ser #status predicted

Query Match 14.9%; Score 513.5; DB 2; Length 381;  
 Best Local Similarity 38.2%; Pred. No. 5.9e-21;  
 Matches 139; Conservative 52; Mismatches 124; Indels 49; Gaps 13;

QY 78 VLRMGAQVYKSYKIIPAVAVKIKARDLLIAGMIDTGYFQNTRVSGIKFIQEDYKQVD 137  
 Db 58 VISERKGVQKQFYVNAATAATLDEKAVKEL-----KKDPVAVYEDH-IAHE 105  
 QY 138 DATSV---SOIGADVTYNSLGYDGSQVVAIVDTGIDANHPDLKGVIGVDMVYVNRST 193  
 Db 106 YAGVPPVPGISQIKPAL-HSGGTGYSNVKVAVIDSGIDSSHPDL--NVRGASFPSETN 162  
 QY 194 PYDD-QGHGTHVAGIVAGTGSVNSQYIGVAPGALVGVKVLGADSGSVSTIAGVDMV 252  
 Db 163 FYQDSSHGTHVAGTIALNN-sIGVLGVAPSAFLVAVKVLDSGTSGSQYSMIINGIEMAI 221  
 QY 253 QNKQKIGRVINILSGSSQSDGTDLSQAVNNAMDAGIYVCAAGNSGP--NTYTVGSP 310

Db 222 SN----NMDVNNLSG---GPTGSLTKTVNDAVSSGIVVAAAAGNESSGSGSTVGY 274  
 Qy 311 AAASKVTYGAVDNSNDNIAFSRSGPTADGRLEKPEVAPGVDDIAPRAGSTSMGTPINDY 370  
 Db 275 AKYPSITIAVGAVNSSQORASFSSVSELD-----VMAPEVSIQSTLPQGT----- 319  
 Qy 371 YTKASGTSMATPVNSGVALIIOAHPSWTPDKYKTLIETADIVAPKEIADIAYGAGRNV 430  
 Db 320 YGAVNGTSMATPVNAGAAALLISKHPTWTAQVRDLREISTATYLG-----NSFYGGGLIN 375  
 Qy 431 VYKA 434  
 Db 376 VQAA 379

RESULT 12  
 SUBST  
 N:Alternates: Bacillus subtilis  
 C:Spec: Bacillus subtilis  
 C:Date: 04-Dec-1986 #sequence revision 04-Dec-1986 #next change 05-Oct-2004  
 C:Accession: A00972; A26116; I39778; I39779; S68012; H6586  
 R:Stahl, M.L.; Ferrari, E.  
 J. Bacteriol. 158, 411-418, 1984  
 A>Title: Replacement of the Bacillus subtilis subtilisin structural gene with an in vitro  
 A:Reference number: A00972; PMID:84212198; PMID:6427178  
 A:Accession: A00972  
 A:Molecule type: DNA  
 A:Residues: 1-381 <STA>  
 A:Cross-references: UNIPROT:P04189; UNIPARC:UPI0000142525; GB:K01988; NID:g143519; PIDN:  
 R:Mong, S.L.; Price, C.W.; Goldfarb, D.S.; Doi, R.H.  
 Proc. Natl. Acad. Sci. U.S.A. 81, 1184-1188, 1984  
 A>Title: The subtilisin E gene of Bacillus subtilis is transcribed from a sigma37 promoter  
 A:Reference number: A26116; PMID:84144862; PMID:6322190  
 A:Molecule type: DNA  
 A:Residues: 1-155 <MON>  
 A:Cross-references: UNIPARC:UPI000015E997; GB:K01443; NID:g143665; PIDN:AAA22814.1; PID:  
 R:Ikemura, H.; Takagi, H.; Inouye, M.  
 J. Biol. Chem. 262, 7859-7864, 1987  
 A>Title: Requirement of pro-sequence for the production of active subtilisin E in Bacche  
 A:Reference number: I39969; PMID:87222417; PMID:3108260  
 A:Accession: I39970  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-156 <IKK>  
 A:Cross-references: UNIPARC:UPI000016E956; GB:M16639; NID:g143521; PIDN:AAA22744.1; PID:  
 R:Renner, D.J.; Ferrari, E.; Perego, M.; Hoch, J.A.  
 J. Bacteriol. 170, 296-300, 1988  
 A>Title: Location of the targets of the hpr-97, sacU32(Hy), and sacQ36(Hy) mutations in  
 A:Reference number: I39778; PMID:88086885; PMID:2447063  
 A:Accession: I39778  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 18 <HEN>  
 A:Cross-references: UNIPARC:UPI000016E7CE; GB:M19125; NID:g142527; PIDN:AAA22245.1; PID:  
 R:Park, S.  
 J. Bacteriol. 171, 2657-2665, 1989  
 A>Title: Bacillus subtilis subtilisin gene (aprE) is expressed from a sigma-A (sigma-43)  
 A:Reference number: I39779; PMID:89213955; PMID:2496113  
 A:Accession: I39779  
 A:Molecule type: DNA  
 A:Residues: 1-13 <PAR>  
 A:Cross-references: UNIPARC:UPI000016E7CE; GB:M1060; NID:g142529; PIDN:AAA22246.1; PID:  
 A:Experimental source: strain W168, substrain PY79  
 R:Kamal, M.; Hoozee, J.O.; Kalsner, R.; Shafgat, J.; Razzaki, T.; Zaidi, Z.H.; Joernvall,  
 P. Lect. 374, 363-366, 1995  
 A>Title: Isolation, characterization and structure of subtilisin from a thermostable Bac  
 A:Reference number: S68012; PMID:96069545; PMID:7589571  
 A:Accession: S68012  
 A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 113-323 <KAM>  
 A:Cross-references: UNIPARC:UPI0000172C13  
 R:Kunze, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertec  
 C.; Bron, S.; Brouillet, S.; Bruchli, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi  
 A.; Enllich, S.D.; Emerson, P.T.; Enllich, K.D.; Erttingen, U.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fume, S.; Gallizi, A.; Gallier  
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;  
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
 A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,  
 Y.; M.; Ogawa, K.; Ogawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
 A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero,  
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zumestein, E.; Yoshikawa, H.; Danchin, A.  
 A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; PMID:98044033; PMID:9384377  
 A:Accession: H69586  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-26, 'V', 28-381 <KUN>  
 A:Cross-references: UNIPARC:UPI00006019A; GB:Z99109; GB:AL009126; NID:g2633260; PIDN:CAF  
 A:Experimental source: strain 168  
 C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and many  
 not necessary for normal sporulation.  
 C:Genetics:  
 A:Gene: aprE  
 A:Map position: 690-771  
 A:Start codon: GTG  
 C:Function:  
 A:Description: catalyzes the hydrolysis of peptide bonds  
 A:Note: this enzyme has broad specificity and will hydrolyze peptide amides; it prefers C  
 C:Superfamily: Subtilisin; subtilisin homology  
 C:Keywords: extracellular protein; hydrolase; protein digestion; serine proteinase  
 F:1-23/Domain: signal sequence #status predicted <SIG>  
 F:24-106/Domain: activation peptide #status predicted <APT>  
 F:107-381/Product: subtilisin E #status predicted <MPT>  
 F:129-341/Domain: subtilisin homology <SBT>  
 F:138, 170, 327/Active site: Asp, His, Ser #status predicted

Query Match 14.9%; Score 512.5; DB 1; Length 381;  
 Best local similarity 37.9%; Pred. No. 6, 7e-21;  
 Matches 138; Conservative 53; Mismatches 124; Indels 49; Gaps 13;

Qy 78 VRLMGAQVYKSYKIIPAVAVKIKARDLLIAGMIDTGYFGNTRVSGIKFIQEDYQVVD 137  
 Db 58 VISEKGVKQKQKVNAAAATLDEKAVEL-----KKDPSVAVEEDH-IAHE 105  
 Qy 138 DATSV-----SOICADTVNMSLGYDGSVVVAIVDTGIDANHPDKKRVGMYAVNGRST 193  
 Db 106 YQASVYVGIQIKAPL-HSQGYTSNVAVVIDSGIDSSHPL--NVKGAASFVSETN 162  
 Qy 194 PYDD-QGHGTHVAGIAGTGVNSQYIVGAPKALVGVKADGSGSVSTIAGVDMV 252  
 Db 163 PYDDSGSHGTHAGTIALAN-SIGVLGVSFPAASLVAIVKLBSTGSGQYSWIINGEMAI 221  
 Qy 253 QNKDKYRIVNLSLSSQSSSDGTDLSQAVNNAMAGIIVCVAAAGNSP--NTYTVGSP 310  
 Db 222 SN----NMDVNNLSG---GPTGSLTKTVNDAVSSGIVVAAAAGNESSGSGSTVGY 274  
 Qy 311 AAASKVTYGAVDNSNDNIAFSRSGPTADGRLEKPEVAPGVDDIAPRAGSTSMGTPINDY 370  
 Db 275 AKYPSITIAVGAVNSSQORASFSSVSELD-----VMAPEVSIQSTLPQGT----- 319  
 Qy 371 YTKASGTSMATPVNSGVALIIOAHPSWTPDKYKTLIETADIVAPKEIADIAYGAGRNV 430  
 Db 320 YGAVNGTSMATPVNAGAAALLISKHPTWTAQVRDLREISTATYLG-----NSFYGGGLIN 375  
 Qy 431 VYKA 434  
 Db 376 VQAA 379











## Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

**When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.**

**This Page Blank (uspto)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2007, 22:20:17 ; Search time 192 Seconds

(without alignments)  
3174.923 Million cell updates/sec

Title: US-10-800-684-1  
Perfect score: 3437  
Sequence: 1 MKRLGAVLALVGLAGT.....YASTYGMADYQLKAVVYG 659

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_7.2.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match | Length | ID           | Description         |
|------------|--------|-------------|--------|--------------|---------------------|
| 1          | 2664   | 77.5        | 663    | 05J125_PYRO  | 05J125 pyrococcus   |
| 2          | 2501.5 | 72.8        | 654    | 08U0C9_PYRPU | 08U0C9 pyrococcus   |
| 3          | 982.5  | 28.6        | 561    | 08R8J2_THETN | 08R8J2 thermoaer    |
| 4          | 937.5  | 27.3        | 424    | 03C8T5_THETN | 03C8T5 thermoaer    |
| 5          | 711    | 20.7        | 795    | 05NM24_9ARCH | 05NM24 uncultured   |
| 6          | 700.5  | 20.4        | 1294   | 050M77_STRSH | 050M77 streptomyc   |
| 7          | 699    | 20.3        | 1245   | 09RL54_STRCO | 09RL54 streptomyc   |
| 8          | 695.5  | 20.2        | 430    | 08RVL1_OCEIH | 08RVL1 oceanobacil  |
| 9          | 674.5  | 19.6        | 1105   | 08KX66_STRVD | 08KX66 streptomyc   |
| 10         | 669.5  | 19.5        | 442    | 031788_BACSU | 031788 bacillus su  |
| 11         | 668.5  | 19.5        | 1139   | 082139_STRAM | 082139 streptomyc   |
| 12         | 662    | 19.3        | 412    | 03C8T3_THETN | 03C8T3 thermoaer    |
| 13         | 657.5  | 19.1        | 1208   | 082814_STRAM | 082814 streptomyc   |
| 14         | 651    | 18.9        | 644    | 046C21_METBA | 046C21 methanosaer  |
| 15         | 649.5  | 18.9        | 442    | 05J135_GEOXA | 05J135 geobacillus  |
| 16         | 649    | 18.9        | 442    | 0651P4_BACID | 0651P4 bacillus ii  |
| 17         | 637    | 18.5        | 444    | 09KJ17_BACHD | 09KJ17 bacillus ha  |
| 18         | 636    | 18.5        | 412    | 08RC68_THETN | 08RC68 thermoaer    |
| 19         | 636    | 18.5        | 1102   | 09S684_STRAO | 09S684 streptomyc   |
| 20         | 633    | 18.4        | 412    | 09AER6_THETN | 09AER6 thermoaer    |
| 21         | 632.5  | 18.4        | 1237   | 08GCT4_STRAZ | 08GCT4 streptomyc   |
| 22         | 626    | 18.2        | 435    | 08EMJ3_OCEIH | 08EMJ3 oceanobacil  |
| 23         | 623    | 18.1        | 1220   | 09L0A0_STRCO | 09L0A0 streptomyc   |
| 24         | 619.5  | 18.0        | 1253   | 09FC06_STRCO | 09FC06 streptomyc   |
| 25         | 616    | 17.9        | 369    | 03CS23_9CLOT | 03CS23 alkaliophilu |
| 26         | 590    | 17.2        | 1239   | 09F8Z4_STRCO | 09F8Z4 streptomyc   |
| 27         | 564.5  | 16.4        | 379    | 09P8P4_BACDI | 09P8P4 bacillus ii  |
| 28         | 564.5  | 16.4        | 379    | 0651P7_BACDI | 0651P7 bacillus ii  |
| 29         | 563.5  | 16.4        | 374    | 09P942_BACDI | 09P942 bacillus ii  |
| 30         | 563.5  | 16.4        | 379    | 06BCN9_BACDI | 06BCN9 bacillus mo  |
| 31         | 563.5  | 16.4        | 379    | 06PNN5_BACDI | 06PNN5 bacillus ii  |

|    |       |      |      |   |              |                    |
|----|-------|------|------|---|--------------|--------------------|
| 32 | 562.5 | 16.4 | 374  | 2 | 09P943_BACDI | 09P943 bacillus ii |
| 33 | 562.5 | 16.4 | 379  | 2 | 04PKR6_BACDI | 04PKR6 bacillus ii |
| 34 | 559.5 | 16.3 | 379  | 2 | 05J521_BACDI | 05J521 bacillus ii |
| 35 | 558.5 | 16.2 | 374  | 2 | 09P941_BACDI | 09P941 bacillus ii |
| 36 | 557.5 | 16.2 | 379  | 1 | SUBT_BACDI   | P00780 bacillus ii |
| 37 | 556.5 | 16.2 | 379  | 2 | 045259_BACDI | 045259 bacillus ii |
| 38 | 553   | 16.1 | 453  | 2 | 03X3G0_9ACTN | 03X3G0 rubrobacter |
| 39 | 550.5 | 16.0 | 379  | 2 | 045300_BACDI | 045300 bacillus ii |
| 40 | 549.5 | 16.0 | 1407 | 2 | 072YJ0_BACDI | 072YJ0 bacillus ce |
| 41 | 549   | 16.0 | 379  | 2 | 045301_BACDI | 045301 bacillus ii |
| 42 | 549   | 16.0 | 1358 | 2 | 08RTM4_OCEIH | 08RTM4 oceanobacil |
| 43 | 547.5 | 15.9 | 514  | 2 | 04H6A0_9DEIO | 04H6A0 deinococcus |
| 44 | 547   | 15.9 | 382  | 2 | 045522_9BACI | 045522 bacillus sp |
| 45 | 544.5 | 15.8 | 376  | 2 | 076KL9_9BACI | 076KL9 bacillus sp |

## ALIGNMENTS

|                       |  |                           |
|-----------------------|--|---------------------------|
| RESULT 1              | 05J125_PYRO  | PRELIMINARY; PRT; 663 AA. |
| ID                    | 05J125_PYRO  |                           |
| AC                    | 05J125_PYRO  |                           |
| DT                    | 15-FEB-2005, integrated into UniProtKB/TrEMBL.   |                           |
| DT                    | 15-FEB-2005, sequence version 1.   |                           |
| DT                    | 07-FEB-2006, entry version 7.  |                           |
| DE                    | Subtilisin-like serine protease.   |                           |
| GN                    | OrdereddictusNames=TKI689;   |                           |
| OS                    | Pyrococcus kodakarensis (Thermococcus kodakarensis).   |                           |
| OC                    | Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  |                           |
| OC                    | Thermococcus.  |                           |
| OX                    | NCBI_TaxID=69014;  |                           |
| RN                    | [1]  |                           |
| RP                    | NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].   |                           |
| RC                    | STRAIN=KOD1;   |                           |
| RX                    | Published=15710748; DOI=10.1101/gr.3003105;  |                           |
| RA                    | Fukui T., Atomi H., Kanai T., Matsumi R., Fujiwara S., Imanaka T.;   |                           |
| RT                    | "Complete genome sequence of the hyperthermophilic archaeon  |                           |
| RT                    | Thermococcus kodakarensis KOD1 and comparison with Pyrococcus  |                           |
| RT                    | genomes.";   |                           |
| RL                    | Genome Res. 15:352-363(2005).  |                           |
| CC                    | Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a> |                           |
| CC                    | Distributed under the Creative Commons Attribution-NonCommercial   |                           |
| CC                    | License.   |                           |
| DR                    | EMBL; AP006878; BAD85878.1; -; Genomic DNA.  |                           |
| DR                    | GO; GO:0008233; F:peptidase activity; IEA.   |                           |
| DR                    | GO; GO:0004289; F:subtilisin activity; IEA.  |                           |
| DR                    | GO; GO:0006508; P:proteolysis; IEA.  |                           |
| DR                    | InterPro; IPR007280; Pept arc bac_C.   |                           |
| DR                    | InterPro; IPR002029; Pept arc S53.   |                           |
| DR                    | Pfam; PF00082; Peptidase_S8; 1.  |                           |
| DR                    | Pfam; PF04151; PC; 2.  |                           |
| DR                    | PRINTS; PR00723; SUBTILISIN.   |                           |
| DR                    | PROSITE; PS00136; SUBTILASE ASP; UNKNOWN_1.  |                           |
| DR                    | PROSITE; PS00137; SUBTILASE HIS; UNKNOWN_1.  |                           |
| DR                    | PROSITE; PS00138; SUBTILASE_SER; 1.  |                           |
| KW                    | Complete proteome; Protease.   |                           |
| SQ                    | SEQUENCE 663 AA; 70955 MW; 2C68ACD3888E90E CRC64;  |                           |
| Query Match           | 77.5%; Score 2664; DB 2; Length 663;   |                           |
| Best Local Similarity | 75.9%; Pred. No. 5.1e-136;   |                           |
| Matches               | 505; Conservative 64; Mismatches 88; Indels 8; Gaps 4;   |                           |
| QY                    | 1 MKRLGAVLALVGLAGTALAAYKPYVRNNVAQKKYGLLTGGLPKYVRNNMNOEV 60   |                           |
| DB                    | 1 MKKFAVVLALFVLGIMGASVLAAPQKAVR-NVSGQKRYGLLTGGLPKYVRMSWDEV 59  |                           |
| QY                    | 61 DIVINEGSGDRDRAVKVLRIMGAQVKYSYKIIIPAVVYKIKARDLLTAGMIDTGYFNT 120  |                           |
| DB                    | 60 STIIFEDNADKEKAVVILDFGAKIKRYNHIIPALAVKIKKIDLLTAGMIDTGYFNA 119  |                           |
| QY                    | 121 RVSGIKFIQEDYKQVDDAT-----SVSGIGADTVNSLGYDGSVVVAIVDTGIDANHP 175  |                           |

```

Db 120 QLSGVQFIQEDYVAVAVETSLDSSAQAQVATMMNM-LGSDSGSITITIGIDGIDASHP 178
QY 176 DLKGVIVMAYAVNRSTPYDDQGHGTHVAGIVAGTGSV-NSQYIGVAPGAKLVGVKLG 234
Db 179 DLQGVIVMAYVFNKTPYDDNGHGTIVASIAAGTGAASNGKYGMAFGAKLVGIVKLN 238
QY 235 ADGSGSVSTIIAGVDMVQNKDKYGIKIVINLSLSSGSSSDGTDLSQAANNAMAGIYVC 294
Db 239 GGGSGSISDIIINGVDMVQNKDKYGIKIVINLSLSSGSSSDGTDLSQAANNAMAGIYVC 298
QY 295 VAAAGNSGNTTVGSPAAASKYITVGAVDSDNMTASFSSRGPTADGRLKEPVAPGVDI 354
Db 299 VAAAGNSGNTTVGSPAAASKYITVGAVDSDNMTASFSSRGPTADGRLKEPVAPGVMI 358
QY 355 APRAAGTSMGPINDYITAAFGTSMATPHVAGIALLLQAHPSWTPKVKITALIETADIY 414
Db 359 AARASGTSMGCPINDYITAAFGTSMATPHVAGIALLLQAHPSWTPKVKITALIETADIY 418
QY 415 APKEIADIAYAGRVNRYKAIKYDDYAKLTFTGSVADKGSATHTFDVSGATFTVATLYMD 474
Db 419 KPDEIADIAYAGRVNRYKAIKYDDYAKLTFTGSVADKGSATHTFDVSGATFTVATLYMD 478
QY 475 TGSSDIDLILYDENGNEVDISYIAYGFEKVGYNPTAGTWTVAVSYKGAANYQVDVVS 534
Db 479 NSGSDILLYLDYDENGNEVDISYIAYGFEKVGYNPTAGTWTIYKVSYSGSANYQVDVVS 538
QY 535 DGSLSQSGGPNPNPNPTPTDIOFTGSVNDVMDTSDTFPMNNSGATKITGLTIF 594
Db 539 DGSLSQSGGPNPNPNPTPTDIOFTGSVNDVMDTSDTFPMNNSGATKITGLTIF 598
QY 595 DTSYNDIDLILYDENGNEVDISYIAYGFEKVGYNPTAGTWTFLVAVSYTGMAADYOLKA 654
Db 599 DTSYNDIDLILYDENGNEVDISYIAYGFEKVGYNPTAGTWTFLVAVSYTGMAADYOLKA 658
QY 655 VVYVYG 659
Db 659 KVVYVYG 663

```

RESULT 2  
Q8UOC9\_PYRFU PRELIMINARY; PRT; 654 AA.  
AC Q8UOC9,  
DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.  
DT 07-FEB-2006, entry version 1.  
DE Alkaline serine protease.  
GN Ordered locus names=PF1670;  
OS Pyrococcus furiosus.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=2261;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;  
RA Weiss R.B., Dunn D.M., Robb P.T., Brown J.R.;  
RT "The complete sequence of the Pyrococcus furiosus genome.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonDerivs License  
CC EMBL; AF010265; AA181794.1; -; genomic\_DNA.  
DR HSSP; Q99405; IMPT.  
DR BioCyc; PFUR186497; PF1670-MONOMER; -;  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:000508; F:protease activity; IEA.  
DR InterPro; IPR007280; Pept\_anc\_bac\_C.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF04151; Ppc; 2.

```

DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR Complete proteome.
SQ SEQUENCE 654 AA; 70231 MW; 1CB145A5F505DB34 CRC64;

Query Match 72.8%; Score 2501.5; DB 2; Length 654;
Best Local Similarity 72.8%; Pred. No. 3; 2e-127;
Matches 484; Conservative 69; Mismatches 93; Indels 21; Gaps 8;

QY 1 MKRLGAVLALVYGLAGTALAAAPKPY--VNNNAVQKNGILRGLFKYQORNNMQ 58
Db 1 MKRLKALILVILVGLVGSVAAPKRYEQVNN--VEKNGILRGLFKYQKLNPNB 57
QY 59 EVDVIMPGSYGDRDAVAVKRLMGAQVYKYLIPAVAVKIKAQDLLIAGNIDTGYRG 118
Db 58 EISTVIVFENHREKEIAVVLIELMGAQVYVYHIIIPALAADLKVRDLVLSGL--TG--G 113
QY 119 NTRVSGIKRFOEDYKQVQVDA-----TSVQIGADLVNMSLGIDSGVVAIVDTGIDAN 173
Db 114 KALDSGVRFQEDYKVTVAELLEGIDESAAQVATVVMN-LGYDSGRTIGIIGIDGIDAS 172
QY 174 HPDLKGVIGWYDAVNGRSTPYDDQGHGTHVAGIVAGTGSV-NSQYIGVAPGAKLVGVK 232
Db 173 HPDLQKGVIGWYDFVNGRSTPYDDHGHGTHVASIAAGTGAASNGKYGMAFGAKLVGIV 232
QY 233 LGADSGSVSTIIAGVDMVQNKDKYGIKIVINLSLSSGSSSDGTDLSQAANNAMAGIY 292
Db 233 LGADSGSVSTIIAGVDMVQNKDKYGIKIVINLSLSSGSSSDGTDLSQAANNAMAGIY 292
QY 293 VCAAGNSGPNRYTGSPPAAASKVITVGAVDSDNMTASFSSRGPTADGRLKEPVAPGV 352
Db 293 VVAAGNSGPNRYTGSPPAAASKVITVGAVDSDNMTASFSSRGPTADGRLKEPVAPGV 352
QY 353 IIAPRASGTSMGCPINDYITAAFGTSMATPHVAGIALLLQAHPSWTPKVKITALIETAD 412
Db 353 IIAPRASGTSMGCPINDYITAAFGTSMATPHVAGIALLLQAHPSWTPKVKITALIETAD 412
QY 413 IYAPKEIADIAYAGRVNRYKAIKYDDYAKLTFTGSVADKGSATHTFDVSGATFTVATLY 472
Db 413 IYAPKEIADIAYAGRVNRYKAIKYDDYAKLTFTGSVADKGSATHTFDVSGATFTVATLY 472
QY 473 WPTGSSDIDLILYDENGNEVDISYIAYGFEKVGYNPTAGTWTFLVAVSYTGMAADYOL 532
Db 473 WPTGSSDIDLILYDENGNEVDISYIAYGFEKVGYNPTAGTWTFLVAVSYTGMAADYOL 532
QY 533 VSDGSLSGGPNPNPNPTPTDIOFTGSVNDVMDTSDTFPMNNSGATKITGLD 592
Db 533 VSDGSLSGGPNPNPNPTPTDIOFTGSVNDVMDTSDTFPMNNSGATKITGLD 592
QY 593 TFDTSYNDIDLILYDENGNEVDISYIAYGFEKVGYNPTAGTWTFLVAVSYTGMAADYOL 652
Db 593 TFDTSYNDIDLILYDENGNEVDISYIAYGFEKVGYNPTAGTWTFLVAVSYTGMAADYOL 652
QY 653 KAVVYVYG 659
Db 648 TAKVYVYG 654

```

RESULT 3  
Q8RBJ2\_THETN PRELIMINARY; PRT; 561 AA.  
AC Q8RBJ2,  
DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.  
DT 07-FEB-2006, entry version 1.  
DE Subtilisin-like serine protease.  
GN Name=AprE2; Ordered locus names=TRB0824;  
OS Thermobacterium tengcongensis.  
OC Bacteria; Firmicutes; Clostridia; Thermobacteriales;  
OC Thermobacteriaceae; Thermobacter.  
OX NCBI\_TaxID=119072;

CC (1)  
CC NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonCommercial license  
CC  
CC EMBL: AE013049: AAM24081.1; -; Genomic DNA.  
CC HSSP: P00782: 25RT.  
CC Biocyc: TTEN119072: TTE0824-MONOMER.  
CC DR GO: 0008233; F: peptidase activity; IEA.  
CC DR GO: 00042802; F: protein self binding; IEA.  
CC DR GO: 0004289; F: subtilase activity; IEA.  
CC DR GO: 0004086; P: negative regulation of enzyme activity; IEA.  
CC DR GO: 0006508; P: proteolysis; IEA.  
CC DR InterPro: IPR002048; Pfam: Pfam\_Ca\_Bd.  
CC DR InterPro: IPR007280; Pfam: Pfam\_Ca\_Bd.  
CC DR InterPro: IPR010259; Pfam: Pfam\_S8.  
CC DR Pfam: P00082; Peptide\_S8; I.  
CC DR Pfam: P00451; PFC: 1.  
CC DR Pfam: P005922; Subtilisin\_N; 1.  
CC DR PRINTS: PR00723; SUBTILISIN.  
CC DR PROSITE: PS00018; EF\_HAND\_1; UNKNOWN\_1.  
CC DR PROSITE: PS00136; SUBTILASE\_ASP; 1.  
CC DR PROSITE: PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
CC DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
CC DR Complete proteome; Hydrolase; Protease; Serine protease.  
CC KW SEQUENCE 561 AA; 5969 MW; BA9C5C2F7083A18 CRC64;  
SQ  
Query Match 28.6%; Score 982.5; DB 2; Length 561;  
Best Local Similarity 40.1%; Pred. No. 5.2e-45;  
Matches 232; Conservative 90; Mismatches 187; Indels 69; Gaps 15;  
QY 3 RLGAVALALVL-VGLAGTALAAPVPRVVRNNAVQOKNYGLTLPGLFKYVORNNM-- 57  
DB 5 QLAKLILSLALILSLISLNEILVQAQPNQINLPIDPS--KIYPSLPQKISMQDSKNKI 62  
QY 58 -----QEVDTVIMFGS-YGRDRAVKYRLMGACQYKSYKIIPAAVAKIRA 102  
DB 63 PDDLLEORLINKPDESEFPVILITFNKEVSDADIFTIAKNIGKFKIRKYKIIPSIANLTR 122  
QY 103 RDLILLIAGMIDTGYFENTEVSGIKFIOEDYKVQVDDATSVSQIGADTVWNSIGYDSGVV 162  
DB 123 SCQINVL-----SKLEIYKQLEYDEBPYATLDTATKMFGITKANSDFGVTKNIT 171  
QY 163 VALIVDTGIDANHPDLK-GKVIQWYDAVNGRSTPYDDOCHGTAVAGIAGTGSVNSQYIGV 221  
DB 172 IALIDTIGIDGNHVDLSGKILGKMDKINKKTPPYDNGHGTAVASIAAGTGANSTYKGV 231  
QY 222 APGAKIVGVKVLGADSGSVSTIIAGVWVQNKRYGIRVYNLSIGSSQSDGTDLSQ 281  
DB 232 APDALIVGIKVLIDANGSGMSTVTAGIDMAVQNKDYGIKVINLSIGTSTSSPDGTDSTSL 291  
QY 282 AVNNMADAGIVCVAGNSGPNVTYVSGPAAASKVITVGAV-----DSNNINIASFSGRPT 337  
DB 292 AVNRAVDSGIIVVVAAGNSGPARTYTGSPAAEKALITVAMADVGLGFLIASFSRGP 351  
QY 338 ADGRLEPVAVPQVDIIAPRASGTSKGTPIINDYTKASGTSNATPHVSGVALLIQAPHS 397  
DB 352 ADGRIRKPIAAPGYNITAAKANS-----VNGYVT-YSGISNATPVAAGTVALMLNAN 404  
QY 398 WTPDKYKTLILIEADIVAPKEIADIYAGRVVNYQAIKTDYAKLTFPGSVAD----- 451  
DB 405 LTPNDAKNIMSTYAKSWGPPS-KNVDYAGAGRLDGYEAIKRYA---NFRGNMINDVBNHY 459  
QY 452 -----KGSATHTPDVSGATF-VTATLY---WDTGSSDIDLXYLDPNGENVDVSYTAY 499

DB 460 ISGYLPGSRYSPTWTFNATNTSYPIATLIIIPDMANYPDFDIYLDPSGLIK-SSTGT 518  
QY 500 YGFEKVGYNYPAGTWTWKVSYKGAANYQVNVDSGS 537  
DB 519 QRQETITLPSQGTGYVAVKVSYSKSGNIFPDLASGS 556  
RESULT 4  
ID Q3CJ85.THEBT PRELIMINARY; PRT; 424 AA.  
AC Q3CJ85;  
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.  
DT 22-NOV-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Peptidase S8 and S53, subtilisin, kexin, sedolisin: peptidase, archaeal  
DE and bacterial C-terminal.  
GN ORFNames=Telh39DRAFT\_0431;  
OS Thermomanaerobacter ethanolicus ATCC 33223.  
OC Bacteria; Firmicutes; Clostridia; Thermomanaerobacteriales;  
OC Thermomanaerobacteriaceae; Thermomanaerobacter.  
OX NCBI\_Taxid=340099;  
RN (1)  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ATCC 33223;  
RG US DOE Joint Genome Institute (JGI-PGF);  
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,  
RA Hammon N., Istrati S., Pitluck S., Richardson P., Richardson P.,  
RT "Sequencing of the draft genome and assembly of Thermomanaerobacter  
ethanolicus 39E."  
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ATCC 33223;  
RG US DOE Joint Genome Institute (JGI-ORNL);  
RA Larimer F., Land M.;  
RT "Annotation of the draft genome of Thermomanaerobacter ethanolicus  
39E."  
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonCommercial license  
CC  
CC EMBL: AAKQ01000003; BA065215.1; -; Genomic DNA.  
CC DR GO: 0008233; F: peptidase activity; IEA.  
CC DR GO: 0004289; F: subtilase activity; IEA.  
CC DR GO: 0006508; P: proteolysis; IEA.  
CC KW Hydrolyase; Protease; Serine protease.  
CC SQ SEQUENCE 424 AA; 44126 MW; 335746D47009B928 CRC64;  
Query Match 27.3%; Score 937.5; DB 2; Length 424;  
Best Local Similarity 49.8%; Pred. No. 1e-42;  
Matches 209; Conservative 58; Mismatches 118; Indels 35; Gaps 10;  
QY 145 IADTVWNSLGYDGGVVAIVDTGIDANHPDLK-GKVIQWYDAVNGRSTPYDDOCHGT 203  
DB 13 VTGDRDGNASYSKNDIVIAVIDTIGDSHVDLAGKVIQWODPVNGKSTPYDDNCHGT 72  
QY 204 VAGIYAGTGSVNSQYITGVAAGKLYGVKVLGADSGSVSTIIAGVWVQNKRYGIRYI 263  
DB 73 VASIAAGTGTGNLTKRGVAPGALVGIKVLDSNGSTMSVTAGIDMAVQNKDYGIKVI 132  
QY 264 NLSLSSQSDGTDLSQAVNNAMADAGIVCVAGNSGPNVTYVSGPAAASKVITVGAV- 322  
DB 133 NLSLSTSSDGTDTSLAVNAGVDSGIIVVVAAGNSGPARTYTGSPAAEKALITVAAVA 192  
QY 323 ---DSNNINIASFSGRPTADGRLKPEVAVPQVDIIAPRASGTSKGTPIINDYTKASGTM 379  
DB 193 DVGEIGFLIASFSRGPADGRIRKPIAAPGYNITAAKANS-----INGYVT-YSGISM 245

```

QY 360 ATPHVSQVAILLOAHPSWTPDKVTALITETADIVAPKEIADIVAGAGRVVYKAIKDD 439
DB 246 ATPFVAGTVALMLSANINILAPDAKNIIMTTAKSMGPS-KNIDYGVGRDLAYEAIKTAG 304
QY 440 -----YAKLFTGSGVADKSGATITFDVSGATFTATLY----WDTGSSDIDL 482
DB 305 NFTGNINISVPHNYAKESLPDS---RYSIDIWTFNTDTSYPIATFTIIPDMANPNPDI 361
QY 483 YLYDPNGNEVDVSYTAYAYGFEKGVYNNPTAGTMTVTVSYKGAANYQVDV-YSDGSLSSQ 541
DB 362 YLYDPTGLVK-SSGTGTGROETITITITPTGTITTIKYSFSGSNYTFDLSVGGGSLTLS 420

RESULT 5
Q5NM24_9ARCH PRELIMINARY; PRT; 795 AA.
ID Q5NM24_9ARCH
AC Q5NM24_9ARCH
DT 04-JUN-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 1.
DE Alkaline serine protease.
GN ORFNames=orf17;
OS uncultured archaeon.
OC Archaea; environmental samples.
OX NCBI_TaxID=115547;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=16329940; DOI=10.1016/j.femsec.2004.12.004;
RA Etkel C., Kemnitz D., Kube M., Ricke P., Chin K.-J., Dedysh S.,
RA Reinhardt R., Conrad R., Liesack W.;
RT "Retrieval of first genome data for rice cluster I methanogens by a
RT combination of cultivation and molecular techniques.";
RL FEBS Microbiol. Ecol. 53:187-204(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs License
CC -----
EMBL: CR626856; CA04782.1; -; Genomic DNA.
DR GO: GO:0008233; F:peptidase activity; IEA.
DR GO: GO:0042802; F:protein self binding; IEA.
DR GO: GO:0004289; F:subtilase activity; IEA.
DR GO: GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO: GO:0006508; P:proteolysis; IEA.
DR InterPro: IPR011964; Beta_ptc_ytcn.
DR InterPro: IPR000209; Pept_S8_S53.
DR InterPro: IPR010259; Prot_inh_S8A.
DR Pfam: PF00082; Peptidase_S8; 1.
DR Pfam: PF05922; Subtilisin_N; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR TIGRFAMs: TIGR02276; beta_ptc_ytcn; 3.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1.
KW Protease.
SQ SEQUENCE 795 AA; 82723 MW; 6295368B8D16B8 CRC64;

Query Match 20.7%; Score 711; DB 2; Length 795;
Best Local Similarity 31.6%; Pred. No. 4.1e-30;
Matches 194; Conservative 92; Mismatches 207; Indels 120; Gaps 16;

QY 57 NOEVDVIVFGSGDRDAVKVLRLLMGAVKYSYKIIIPAVAKIKARDLLLAQMIDTCY 116
DB 31 NKGQASLISQGAASGKATRLRVKAGKVKNNYIIDAVALDPDSEVAALAKAPD--- 87
QY 117 FGNTRVSGIKFTQEDYKQVVDATSVSIQADTVNSL-GYDGSQVVAIVDTGIDANHP 175
DB 88 -----VESVVRDSIAVADLDEVVYHNASVANGSTVGYTGKGVNSVIDSGIDAHHP 139
QY 176 DLKGVITGVYDVANGSTPYDQGHGTHVAGIVAGTGSVNSQYT-GVAPGAQLVGVKVLG 234
DB 140 DLAKGVITLMDPLNDATPYDDFGHGTFAAGIIASGASGGEVGLVYDASLFFGVKVLN 199
QY 235 ADGSGSVSTIIAGVDVNVQNKDKYGRIVINILSGSSQSGSDGTDSLQAVNNAMMDAGIYVC 294

```

```

DB 200 PSGTAYVSDIILAAIDSVQNH-----ADVLSMSLSNPTH---IQALDAVHNAENSVVAV 252
QY 295 VAAAGSGPNTYTVGSPAAASKVITVGAVDNSNDVIASFSSRGPTADGRLLKPEVAPGVDTI 354
DB 253 CSAAGNTGPKYKGSIRCGDSDPDVIAVGSVMSBRLSSFSRSGPTQDRIRPDIVAVEYVI 312
QY 355 APPASGTSMTPIINDYTTAASGTSNATPHVSGVAILLOAHPSWTPDKVTALITETADIV 414
DB 313 SYRSSGSGTWGNPNIGQYCYASGTSAACPQVNASASAILLOANGSLTFPEELKDVILIRNT--- 369
QY 415 APRKIIDI-----AYGAGRVVYKAIKDDYAKLFTGSAVDKSGATHTDVGATFTVAT 470
DB 370 --YHLSDTYPENQGGKRLINIAL--NEVLQVTPPTPTPTPTATPTATPTATPTATPTAT 425
QY 471 LYWDGSSDIDLXYDPNGNEVDVSYTAYAYGFEKGVYNNPTA---GTMTVTVSYKGAAN 527
DB 426 -----FLATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 440
QY 528 YQYDVVSDGSLSSQGGGPNPNPNPNPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 570
DB 441 -----TVPTPTATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 491
QY 571 WDTSDTFTNN-----VNSGATKITGD-LTFDTSYNDLXYDPNGNLVDRSTSSNS 621
DB 492 PPKSKTFVNSGNTVSVIDGSTNTVTGIIIGSRPYG-----IYVPSDGLV--YVAVVG 545
QY 622 YEHVEYANAPGT 634
DB 546 LNRVAIISPANNT 558

RESULT 6
Q5OHM7_STRSH
ID Q5OHM7_STRSH PRELIMINARY; PRT; 1294 AA.
AC Q5OHM7;
DT 07-JUN-2005, integrated into UniProtKB/TrEMBL.
DT 07-JUN-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Probable secreted peptidase.
OS Streptomyces sphaeroides.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=195949;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCIMB 11891;
RX PubMed=15870333; DOI=10.1128/AEM.71.5.2452-2459.2005;
RA Bustaquito A.S., Gust B., Galm U., Li S.-M., Chater K.F., Heide L.;
RT "Heterologous Expression of Novobiocin and Chlorobiocin Biosynthetic
RT Gene Clusters.";
RL Appl. Environ. Microbiol. 71:2452-2459(2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCIMB 11891;
RA Steffensky M., Li S.-M., Heide L.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs License
CC -----
EMBL: AY227005; AAP48601.1; -; Genomic DNA.
DR GO: GO:0008233; F:peptidase activity; IEA.
DR GO: GO:0004289; F:subtilase activity; IEA.
DR GO: GO:0006508; P:proteolysis; IEA.
DR InterPro: IPR001337; PA.
DR InterPro: IPR00209; Pept_S8_S53.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 1294 AA; 13576 MW; 804C7F9A0DDEB896 CRC64;

```



Query Match 20.4%; Score 700.5; DB 2; Length 1294;  
 Best Local Similarity 45.6%; Pred. No. 2,8e-29;  
 Matches 160; Conservative 53; Mismatches 117; Indels 21; Gaps 9;

93 IPAAVAVKIKARDL-----LLIAGMIDTGYFGNTRVSGIKFOEDYKQVDDATSVSQIG 146  
 165 IDAAVAVRPAVLGAFWKLAPAG-----PSINSALDAVPKWLDRVGSALDRSTAQIG 220  
 147 ADTWNSLIGYDGSQVVAIVDTGIDANHPDLKGIWYDAVANGRTPTDDGCHTHVAG 206  
 221 AEDVWKS-GLRGERVAVVLTGADQTHPDLGRIMAAED-PSGSGTADGSGHGHVAVS 278  
 207 IYAGSGSVN-SQYIGVAPGAKLVGVVVLGADSGSVSTIIAGVDVWQKDKGIRVINTL 265  
 279 IYGGSGKASGGTRQGAAPAEELMIGKVLDDGDFGSGSYIAGMEMAAK-----GAEVVM 334  
 266 SLGSSQSDGDTSLGQAVNN-AMDAGIVVCAAGNSGPNYTVGSPAAASKYITVGVADS 324  
 335 SLGSDAPSDGTPMSLAVNELSSGALFVVAAGNSPGSGTIGSGAADALITVGAVD 394  
 325 NNIIASFSSRRP-TADGRLEPVAVPGVDIIAPRASGTMGTFINDYTKASGTMATH 383  
 395 DDLAEFSSRGPSGDEAVKPDVTAPGVGIVAARATGTGMDPVDGDTYTAASGTMATH 454  
 384 VSGVALLIQAHPSTPDPVKTALETFADIAPKEIADIAYAGARVNYKA 434  
 455 VAGAAALLAQRHPDWSAAQLKDALVSTARTIAGQVTE--QGGGRIDILAA 503

## RESULT 7

Q9RL54\_STRCO PRELIMINARY; PRT; 1245 AA.

Q9RL54; 01-MAY-2000, integrated into UniProtKB/TrEMBL.

DT 01-MAY-2000, sequence version 1.

DT 07-FEB-2006, entry version 21.

DE Probable secreted peptidase.

GN OrderedLocNames=SC00432; ORFNames=SCF51A.10;

OS Streptomyces coelicolor.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycinae; Streptomycetaceae; Streptomycetes.

NCBI\_TaxId=1902;

[1] NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=A3(2) / M145;

RC MEDLINE=21996410; PubMed=1200953; DOI=10.1038/417141a;

RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,

Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieseer H.,

Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,

Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,

Huang C.-H., Kieseer T., Larke L., Murphy L.D., Oliver K., O'Neill S.,

Rabinovitch E., Rajandream M.A., Ruberford K.M., Ruter S., Taylor K.,

Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,

Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,

Hopwood D.A.;

"Complete genome sequence of the model actinomycete Streptomyces

coelicolor A3(2).";

Nature 417:141-147(2002).

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NonDerivs License

CC EMBL; AL939105; CAB5662.1; -; Genomic\_DNA.

DR HSSP; P00782; 1SUR.

DR BIOCyc; SC0B1902; SC0432-MONOMER; -;

DR GO; GO:0008233; F:peptidase activity; IEA.

DR GO; GO:0006508; F:protease activity; IEA.

DR InterPro; IPR003137; PA.

DR InterPro; IPR000209; Pept\_S8\_S53.

DR InterPro; IPR010221; VCBS-

DR Pfam; PF02225; PA; 1.

DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR TIGRFAMs; TIGR01965; VCBS repeat; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 1245 AA; 130896 MW; 74EE92DB9CA1DE60 CRC64;

Query Match 20.3%; Score 699; DB 2; Length 1245;  
 Best Local Similarity 41.1%; Pred. No. 3.2e-29;  
 Matches 175; Conservative 67; Mismatches 134; Indels 50; Gaps 13;

72 RDAVAVKRLMGAQV-----KSYKIIIPAAVKIKARDLLIAGMIDTGYFGNTRV 122  
 157 RDAVRELDSTDAVAVRDEALGRFMRKLVPAAGQRAKA-----A 197  
 123 SGIRKFOEDYKQVDDATSVSQIGADTWNSLIGYDGSQVVAIVDTGIDANHPDLKGYI 182  
 198 AATPRVWLDGRVGAALDRSTAQIGAPDVW-SAGYRGDGKVAIVLTGADQSHPDLAGRYA 256  
 183 GMYDAVNGRSTPYDQGHVAGIIVAGSVN-SQYIGVAPGAKLVGVVVLGADSGSV 241  
 257 AAKD-PSGSSGNTDVFHGHVASTVGSAGSAGSROGVAAPAAKLVKVLGDDGFSB 315  
 242 STIAGVWVWQNKQRYGIRVINTSLGSSQSDGDTSLGQAVNN-AMDAGIVVCAAGNS 300  
 316 SQVINGMEWAADQ-----GADVNNMISLSSGATDGTDPMSQALNLSRRTGTLFVVAAGNE 371  
 301 G-PNTYVGSPPAAKRVITVGAVDSDNDNIASFSSRGP-TADGRLEPVAVPGVDIIAPRA 358  
 372 GEGQPRVSGSPGAADALITVGAVDSDSDSLAPSSSRGPRGADVAVXPDVAPGVIAAARA 431  
 359 SGTSMGTPINDYTKASGSMATPHVSGVGAIIILAHPSWTPDKVTALETFADIAPKE 418  
 433 AGSAGMDPDEHYTASGTSMTTPHVAAGAAALLAQHPMTGAQLDALISTAVYDQK 491  
 419 IADIVYAGRVVVAIKYDPAKLTFTGSVADKSGATTFPVSGATFVATLTYMDTSS 478  
 492 VTE--QGGGRIVRAA-----GIGAVTATGTL-----VMGPFISRDPEPTSVRY-TNS 539  
 479 DIDVLT 484  
 540 DEDVTL 545

## RESULT 8

Q8ENV1\_OCEIH PRELIMINARY; PRT; 430 AA.

Q8ENV1; 01-MAR-2003, integrated into UniProtKB/TrEMBL.

DT 01-MAR-2003, sequence version 1.

DT 07-FEB-2006, entry version 19.

DE Intracellular alkaline serine proteinase.

GN OrderedLocNames=OB2375;

OS Oceanobacillus ihenyensis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.

NCBI\_TaxId=182710;

[1] NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=HTB831 / DSM 14371 / JCM 11309;

RC MEDLINE=2220767; PubMed=12235376; DOI=10.1093/nar/gkf526;

RA Takami H., Takaki Y., Uchiyama I.,

Genome sequence of Oceanobacillus ihenyensis isolated from the Iheya

Ridge and its unexpected adaptive capabilities to extreme

environments";

Nucleic Acids Res. 30:3927-3935 (2002).

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NonDerivs License

CC EMBL; BA000028; BA014331.1; -; Genomic\_DNA.

DR HSSP; Q99405; IMPT.

DR BIOCyc; OIH8182710; OB2375-MONOMER; -;

DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004289; F:subtilase activity; IEA.  
 DR InterPro: IPR0006508; P:proteolysis; IEA.  
 DR Pfam: PF00082; Peptidase\_S8; 1.  
 DR PRINTS: PR00723; SUBTILASIN.  
 DR PROSITE: PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE: PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
 DR Complete proteome; Hydrolase; Protease; Serine protease.  
 KW SEQUENCE 430 AA; 45839 MW; 6D09A99B8C1310F CRC64;

Query Match 20.2%; Score 695.5; DB 2; Length 430;  
 Best Local Similarity 47.0%; Pred. No. 1.3e-29;  
 Matches 155; Conservative 50; Mismatches 104; Indels 21; Gaps 8;

QY 125 IKFIOEDKVVQVDDATVSOIGADTVNNSLGYDGGVVAIVDTGIDANHPDLKGVIGW 184  
 DB 107 IEKLYDRKVPFLDTHSSINADVLKES-GLTQSGTIAVIDTGIHP-HEDEGRITIGF 164  
 QY 185 YDAVNGSTPYDDQGHGTHVAGIVAGTGSV-NSQYIGVAPQAKLVGVKVLGADSGSVST 243  
 DB 165 ADFYKQGTPEYDNGHGTCHCAGDAAGCALSDQYQGPAPDANLVGVKVLNKTGSSGLST 224  
 QY 244 IIAGVNWNKDKYGRVNLSTGS--SQSSDSTDSLQAVNNAMPAGIVCYAAGNSG 301  
 DB 225 VIBSDICIQKSTKINILSLSGSDATPEAG-DEVVNAVETAMNGVVCVAAAGNSG 283  
 QY 302 PNTYVSGPAAASKVITVGAVDNSN-----DNIAFSRSRPTADGRLEKEVAVPVDII 354  
 DB 284 PGDKTVSGPSIPVITVGAADDNNTARSDDSVAEFSRSRPTIDGLTKRNILTPVDIV 343  
 QY 355 APRASG-----TSMGTINDYTKASGTSMAATPHVSGVALIIQAPSWTPDKVTALIE 409  
 DB 344 SLRPSGFIIDKTNKSAFVGSNYISLSGTSMATPICAGIVQLQSDSLTPNQVREKLM 403  
 QY 410 TADIVAKETIADIVYAGRVNVVYKIKYDD 439  
 DB 404 ACQDLG---SPNVQAGYIMANLIVINE 430

RESULT 9  
 Q8KH6\_STRVD  
 AC Q8KH6\_STRVD PRELIMINARY; PRT; 1105 AA.  
 DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.  
 DT 01-OCT-2002, sequence version 1.  
 DT 07-FEB-2006, entry version 16.  
 DE 1,4-dihydropyridine enantioselective esterase precursor.  
 GN Name-dhpA;  
 OS Streptomyces viridosporus.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=67581;  
 RN [1] \_\_\_\_\_  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=A-914;  
 RX MEDLINE=22034940; PubMed=12039725;  
 RA DOI=10.1128/AEM.68.6.2716-2725.2002;  
 RA Aribea A., Matsufuji M., Nakashima T., Dobaishi K., Ieshiki K.,  
 RA Yoshioka T., Yamada S., Momose H., Taguchi S.;  
 RA "Streptomyces serine protease (DHP-A) as a new biocatalyst capable of  
 RT forming chiral intermediates of 1,4-dihydropyridine calcium  
 RT antagonist.",  
 RL Appl. Environ. Microbiol. 68:2716-2725(2002).

CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
 CC Distributed under the Creative Commons Attribution-NonCommercial License  
 CC EMBL; AB007809; BAC00500.1; -; Genomic\_DNA.  
 DR HSSP; P00782; 2SRT.  
 DR MEROPS; S08.069; -;  
 DR GO; GO:0008233; F:peptidase activity; IEA.

DR GO; GO:0004289; F:subtilase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis; IEA.  
 DR InterPro: IPR000209; Pept\_S8\_S53.  
 DR Pfam: PF00082; Peptidase\_S8; 1.  
 DR PRINTS: PR00723; SUBTILASIN.  
 DR PROSITE: PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE: PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
 DR Hydrolase; Protease; Serine protease; Signal.  
 FT SIGNAL 1 204  
 KW SEQUENCE 1105 AA; 114126 MW; 6AB34D8870579AB CRC64;

Query Match 19.6%; Score 674.5; DB 2; Length 1105;  
 Best Local Similarity 30.8%; Pred. No. 5.8e-28;  
 Matches 219; Conservative 85; Mismatches 273; Indels 133; Gaps 27;

QY 22 LAAPVYVNRNNAVQKATGLTPGLFKVQRMNQEVDTYIMGSGYDRDRAVKTLRL 81  
 DB 97 LPDAARLVAASGLDRRLFDITELG--KAATNSQKQGLKIV--GYGAARAARAAEVR 151  
 QY 82 MGAOVKYSKTIIPAVAVKIKARDLLIAGMIDTVYFNGNRVSGIKFIOEDKVVQVDDATS 141  
 DB 152 EAGELRRITLSINADAVRTPHED--ASRLMDAVTNGDRTASGIAHWLWDGVRRAALDT 208  
 QY 142 VSGIGADTVNNSLGYDGGVVAIVDTGIDANHPDLKGVIGWYDAVNGSTP--YDDQ 199  
 DB 209 VQIGAPKAM-SAGYDGKGVKIAVLDTVGTSHPDLKGVTA--SKNFTAPAGADKVG 264  
 QY 200 HGTNVAIGVAGTGSVN-SQYIGVAPQAKLVGVKVLGADSGSVSTIIAGVDVWNKDKY 258  
 DB 265 HGTNVAISIAAGTGAOSKGVKVPAGAAIINQKVLDDSGFGDSSGIIAGMEWAAAQ---- 320  
 QY 259 GIRVNLSTGSSQSSDGTDSLQAVNN-AMDAIVCYAAGNSGNTYVSGPAAASKVI 317  
 DB 321 GADVNNMSIGMDTPE-IDPLEAIVDKLSAEKGV.PALAAEGEGR-SIGSPGADAAAL 377  
 QY 318 TVGAVDSDNINASFSSRGP-TADGRLEKEVAVPVDIIAPRASGTMGTPIIND--YYTK 373  
 DB 378 TVGAVDSDKXKLADFSSTGPRLDGAIKPDVTA PGVDITAAASBGNDIQEVEGEPAGVMT 437  
 QY 374 ASGTSMAATPHVSGVALIIQAPSWTPDKVTALLETADIVAPKFIADIANAGAVNYK 433  
 DB 438 ISGTSMAATPHVAGAAALIKQHPMTSALKQAL--TGSTKGK-YTFPEQSGSIGIADK 494  
 QY 434 ATK-----YDYAKLFTFGSVAADGSAH 457  
 DB 495 ALQGVIVADPVSVSRVQGWPHHTDEPVTXQLYTNLGTQVTLKLTSTATIDPKKAAPA 554  
 QY 458 TFDVSGAFTVATLTYDTGSSDIDLYLYDPNGNEVDYTAIYGFEEKVGYNPTAGTWT 517  
 DB 555 GFTLGATTVVPA--GGSASVDMTADTRLGTVDGAYSAVYATGGQTVRTAAAVQR 611  
 QY 518 KVSF-----KGAANYQVDVVDGSLSSQSGGNPNPNPNPTPTTDTQTFGS 566  
 DB 612 EVESIDTVVRHIGRDGKFTLHLDLIGVAGISGRTG-----APATDTATRLP 662  
 QY 567 VNDY-----W-----DTSDFTNVNSGATKITGLTLF-DTSYND 600  
 DB 663 KGYLVDSWIAKDRGLKGIQDWLVQPLSTYTKOI-TLTLDAARTK-AAADITVDPRKAP 720  
 QY 601 LDL--YYIDNGNLVDRSTSSNVEHYEVANPAP-----GTWT 636  
 DB 721 LSATIGTYTDTAG--IGIGASMESFADVRMAHLGPAPGIRQTWNGWT 768

RESULT 10  
 C31788\_BACST  
 ID C31788\_BACST PRELIMINARY; PRT; 442 AA.  
 AC C31788;  
 DT 01-JAN-1998, integrated into UniProtKB/TrEMBL.  
 DT 01-JAN-1998, sequence version 1.  
 DT 07-FEB-2006, entry version 31.  
 DE Alkaline serine protease.

GN Name: apr\_1 OrderedLocustNames=BSU17260;  
 OS Bacillus subtilis.  
 CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC SPRAIN-168;  
 RX MEDLINE=96044033; PubMed=9384377; DOI=10.1038/36786;  
 RA Kunzer F., Ogasawara N., Moszer I., Albertin A.M., Alloni G.,  
 RA Azevedo V., Baetero M.G., Bessières P., Bolotin A., Borchert S.,  
 RA Bories R., Boutsier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Bruclet S., Brusch C.V., Caldwell B., Capuano J., Carter N.M.,  
 RA Choi S.-K., Codani J.-J., Comercon I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Duesterhoeft A., Ehrlich S.D., Emmerson P.T.,  
 RA Entian K.-D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Gilm S.-Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Haga K., Haeche J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holappi S., Hosono S., Hulio M.-F., Itaya M.,  
 RA Jones L.-M., Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M.,  
 RA Klein C., Kobayashi Y., Koeltter P., Koningsstein G., Krogh S.,  
 RA Krumm M., Kurita K., Lapidus A., Lachinols S., Laber J.,  
 RA Ladarovic V., Lees S.-M., Levine A., Liu H., Maeda S., Meinel C.,  
 RA Leduc C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S.,  
 RA Medicine C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S.,  
 RA Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B.,  
 RA Park S.-H., Parro V., Pohl T.M., Portetelle D., Potwolik S.,  
 RA Prescott A.M., Prescan E., Pujic P., Purnelle B., Rapoport G.,  
 RA Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B.,  
 RA Rose M., Sadie Y., Sato T., Scanlan E., Schleich S., Schroeter R.,  
 RA Scofield F., Seliguchi J., Sekowska A., Seror S.J., Serron P.,  
 RA Shin B.-S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H.,  
 RA Takemaru K., Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P.,  
 RA Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vanlier F.,  
 RA Vassarotti A., Varti A., Wambutt R., Wedler E., Wedler H.,  
 RA Wetzinger T., Winters P., Wipat A., Yamamoto H., Yamae K.,  
 RA Yasumoto K., Yata K., Yoshida K., Yoshikawa H.-F., Zumbstein E.,  
 RA Yoshikawa H., Zarchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis.";  
 RL Nature 390:249-256 (1997).  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonCommercial License  
 CC  
 CC EMBL/299113; CAB13610.1; -; Genomic\_DNA.  
 DR PIR; A69587; A69587.  
 DR HSP; O99405; IMPR.  
 DR Biocyc; BSUB1423; BSU1727-MONOMER; -;  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004289; F:subtilase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis; IEA.  
 DR InterPro; IPR000209; Pept\_S8\_S53.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SRR; 1.  
 KM Complete proteome; Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 442 AA; 47906 MW; B96446ABE7BMDP2 CRC64;  
 Query Match 19.5%; Score 669.5; DB 2; Length 442;  
 Best Local Similarity 46.0%; Pred. No. 3,6e-28;  
 Matches 151; Conservativity 51; Mismatches 103; Indels 23; Gaps 9  
 Db 116 YLNRVKKLLPLTAEASH-AKEVNVNKGTLTGKGVYVAVDGI-YPHDLGRITIGFAD 173  
 Qy 127 FIOEDYKQVUDANSVSGIQGADYTNWSLIGYDSSGVVAIVPDGIDANPHDKGKXIGWVD 186  
 116 YLNRVKKLLPLTAEASH-AKEVNVNKGTLTGKGVYVAVDGI-YPHDLGRITIGFAD 173  
 Qy 187 AVNGSRSTPDQGHGTHTVAGIYAGTGSVNS-QYIVAVGAKLVGVKGLADSGSGSVTII 245  
 174 MVNQKTEBYDNGHOTHGACGVDASSGSGYRPAPEANLIGVKVLKNGSGSTLADII 233  
 Qy 246 AGVDNVVO-NKDK-YGIRVNLISIGSS--QSSDGTSLSGAVNNMADGIVCVCAAGN 299

| Query                 | Subject  | Score        | DB            | Length |
|-----------------------|--|--------------|---------------|--------|
| Db                    | EGVEMCQYNEEDNDPERIDIMSMSCGPAALRYDHQGEPLVAAVEAMSGAGIVVCAACN   | 23.4         | 293           |        |
| Qy                    | 300 SGPNTYTVGSPAASKVITTVGAVDSN-----DNIAFPSSRGFLPADGRLKEVVAAGVD   | 352          |               |        |
| Db                    | SGPDSQRIASPGVSEKVTITVGLDDNNTPASSDDDDTVASFSSGPIYVYGKEKEDILAPGVN   | 294          | 353           |        |
| Qy                    | 353 IIAARASGT-----SMGTFPINDYTKAGCTSMATPHVSGVALLIIQAHSMTWPKDKTKAL   | 407          |               |        |
| Db                    | 354 IISLRSPSYIDKLQKSSRSVSQYFTWMTSGTSMATPIICAGIALILIQNPDLTPDEVKELL  | 413          |               |        |
| Qy                    | 408 IETADIVAPKEIADIAYGAGGVNYYKAI   | 435          |               |        |
| Db                    | 414 KNGTD--KMKDEDPNITYGAGVNAENSV   | 439          |               |        |
| RESULT 11             |  |              |               |        |
| 082139 STRAW          |  |              |               |        |
| ID                    | Q82139 STRAW   | PRELIMINARY; | PRT: 1139 AA. |        |
| AC                    | Q82139   |              |               |        |
| DT                    | 01-JUN-2003, integrated into UniProtKB/TrEMBL.   |              |               |        |
| DT                    | 01-JUN-2003, sequence version 1.   |              |               |        |
| DT                    | 07-FEB-2006, entry version 16.   |              |               |        |
| DE                    | Putative subtilisin-like protease.   |              |               |        |
| GN                    | OrderedLocustNames=SAV3319;  |              |               |        |
| OS                    | Streptomyces avermectilis.   |              |               |        |
| OC                    | Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;  |              |               |        |
| OC                    | Streptomycinae; Streptomycetaceae; Streptomycetes.   |              |               |        |
| OX                    | NCBI_TaxId=33903;  |              |               |        |
| LN                    | [1]  |              |               |        |
| NP                    | NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].   |              |               |        |
| RC                    | STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;   |              |               |        |
| RC                    | MEDLINE=22608306; PubMed=12692562; DOI=10.1038/nbt820;   |              |               |        |
| RA                    | Ikada H., Ishikawa J., Hanamoto A., Shinoue M., Kikuchi H., Shiba T.,  |              |               |        |
| RA                    | Sakaki Y., Hattori M., Omura S.;   |              |               |        |
| RT                    | "Complete genome sequence and comparative analysis of the industrial   |              |               |        |
| RT                    | microorganism Streptomyces avermectilis."  |              |               |        |
| RL                    | Nat. Biotechnol. 21:526-531(2003).   |              |               |        |
| LN                    | [2]  |              |               |        |
| NP                    | NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].   |              |               |        |
| RC                    | STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;   |              |               |        |
| RC                    | MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211431398;   |              |               |        |
| RA                    | Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  |              |               |        |
| RA                    | Shinoue M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  |              |               |        |
| RA                    | Kikuchi H., Shiba T., Sakaki Y., Hattori M.;   |              |               |        |
| RT                    | "Genome sequence of an industrial microorganism Streptomyces   |              |               |        |
| RT                    | avermectilis: deducing the ability of producing secondary  |              |               |        |
| RT                    | metabolites."  |              |               |        |
| RL                    | Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  |              |               |        |
| CC                    |  |              |               |        |
| CC                    | Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a> |              |               |        |
| CC                    | Distributed under the Creative Commons Attribution-NoDerivs License  |              |               |        |
| CC                    | -----  |              |               |        |
| CC                    | EMBL; BA000030; BAC71030.1; -; Genomic_DNA.  |              |               |        |
| CC                    | HSSP; Q99405; IMPT.  |              |               |        |
| DR                    | MEROPE; S08_069; -.  |              |               |        |
| DR                    | BIOCYC; SAVE227882; SAV3319-MONOMER; -.  |              |               |        |
| DR                    | GO; GO:0008233; F:peptidase activity; IEA.   |              |               |        |
| DR                    | GO; GO:0004289; F:subtilase activity; IEA.   |              |               |        |
| DR                    | GO; GO:0006508; P:proteolysis; IEA.  |              |               |        |
| DR                    | InterPro; IPR002860; Glyco_hydro_BNR.  |              |               |        |
| DR                    | InterPro; IPR000209; Pept_S8_553.  |              |               |        |
| DR                    | Pfam; PF00082; Peptidase_S8; 1.  |              |               |        |
| DR                    | PRINTS; PRO0723; SUBTILISTIN.  |              |               |        |
| DR                    | PROSITE; PS00136; SUBTILASE_ASP; 1.  |              |               |        |
| DR                    | PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.  |              |               |        |
| DR                    | PROSITE; PS00138; SUBTILASE_SSR; 1.  |              |               |        |
| DR                    | Complete proteome; Protease.   |              |               |        |
| Qy                    | SEQUENCE 1139 AA; 118193 MW; EC3B6D234FA94FA0 CRC64;   |              |               |        |
| Query Match           | 19.5%; Score 668.5; DB 2; Length 1139;   |              |               |        |
| Beet Local Similarity | 32.3%; Pred.No. 1.3e-27;   |              |               |        |

Matches 222; Conservative 95; Mismatches 261; Indels 109; Gaps 26;

QY 29 VYRNAVQOKNYGLTPGLFKKYQAMNN-----OEVDVIMFGSYGRDAAVKTLRLMG 83  
 DB 124 VVPADAAAMIAATKGLDQRLF-DVTEINSAIRKAKQGLKYVGKGPASAASAKSDVRAG 182  
 QY 84 AOVKSYKIPAVAVKIKARDLLIAGMIDTGFPGNTFVSGIKFOEDYKQVVDATSVS 143  
 DB 183 T-LRRSLKSLNADAAQTPTD---TAEIMDAVTNGDATASGAHVWLDGTAKSLDSVP 238  
 QY 144 QIGADTVNNSLGGDSGVVAIVDGDIDANHPDLKGYIGYDAVN--GRSTPYDDQHG 201  
 DB 229 QIGAPTAM-AAGYDCKGVKINAVLDGVDATPDLKDQVA---ESKNESAADAADHFGHG 294  
 QY 202 THVAGIVAGTGS-VNSQYIGVAPGAKIVGVKVLGADSGSVSTIAGVDVWYONKDKXI 260  
 DB 295 THVASIAGTAKSGKGYKGVAPGATILNGKVLDDTSGSDSGILLAGEMAAG---GA 350  
 QY 261 RVINLSLGGSSSDGTSLQAVNN-AMDAGIVVCVAAGNSP-NYTVGSPAASRVIT 318  
 DB 351 DVNLSLGGGDTPE-IDPLEAEVVKLSEKGIILPAIAAGNGEFGEGQTIISPGSADALT 409  
 QY 319 VGAVDSNDNIASFSSRGPTADGRLEKPEVAVAGVDIIAPRASGSMGTPIN---DYTKAS 375  
 DB 410 VGAVNDSBDKLASFSSRGGLDGAIKPDTAPGVDTAAAPGSIIDGEGQKPDGYLTIS 469  
 QY 376 GTSMATPHVSGVALIIQAPSWTPDKVKTALLETADIAPKEIADIAGAGRVNYKAI 435  
 DB 470 GTSMATPHVAGAAAILKQHPHNSFAELKALTGA---KQKGYTPQSGSRIADVKA 526  
 QY 436 KYDDYAKLTFGVSADGSAAT-----HTFDVSGAIFVYATL-YMTGSSDIDLVI--- 484  
 DB 527 K-----QSIVANPNVSFGLIQMPHDTDKP---VTQQLRYRLGTSVTLNLAST 573  
 QY 485 -YDENGNEVDVSYTAYVGFKEVGYNPTAGTWTYKVSYKGAANYQDVVSDSLG--- 539  
 DB 574 ATNPKG-----VAAPSGFPLGATKYTVPAGGASADF---TYNTKLGSTTDAYSAVYT 625  
 QY 540 QSGGNNP-----NPNPNPTPTTDTOTFTGVSVDVY---DTG 574  
 DB 626 ATGGQGTVRTAAAVQREVSVDYTLKHIDRDKPAVVSYDLTGVSGLADKFPAPYDAS 685  
 QY 575 DTFPMNVNSGATKLTGLTFD---TSYNDLDLYLDPNGLNLDVRS---SSSYEHVEY 627  
 DB 686 GTYKVRPKNFILNASLFPADPEFTKGAD---WIAQPKLSVTKTIVTADAKAKAPVD 742  
 QY 628 ANPAGTWTFLVYVSYTYGMADYQLKA 654  
 DB 743 TVPDKG-----AKSAFASPDYTVEA 762

RESULT 12  
 O3CCT3 THEET PRELIMINARY; PRT; 412 AA.  
 AC O3CCT3;  
 DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.  
 DT 22-NOV-2005, sequence version 1.  
 DE 07-FEB-2006, entry version 4.  
 DE Peptidase S8 and S53, subtilisin, kexin, sedolisin.  
 GN ORFNames=tech39DRAPT\_1862;  
 OS Thermomicrobacter ethanolicus ATCC 33223.  
 OC Bacteria; Firmicutes; Clostridia; Thermomicrobacteriales;  
 OC Thermomicrobacteriaceae; Thermomicrobacter.  
 NC NCBI\_Taxid=340099;  
 RN NUCLEOTIDE SEQUENCE.  
 RC STRAIN=ATCC 33223;  
 RG US DOE Joint Genome Institute (JGI-PGF);  
 RA Copeland A., Lucas S., Lapidus A., Barry K., Deter J.C., Glavina T.,  
 RA Hammon N., Israni S., Plunk S., Richardson P.,  
 RT "Sequencing of the draft genome and assembly of Thermomicrobacter  
 ethanolicus 39e."  
 RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.

RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=ATCC 33223;  
 RG US DOE Joint Genome Institute (JGI-ORNL);  
 RA Laximer F., Land M.;  
 RT "Annotation of the draft genome of Thermomicrobacter ethanolicus  
 39e."  
 RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 preliminary data.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonCommercial License  
 CC  
 DR EMBL, AA001000019; E064328.1; -; Genomic DNA.  
 DR GO:GO:0008233; F:peptidase activity; IEA.  
 DR GO:GO:0004289; F:subtilisin activity; IEA.  
 DR GO:GO:0006508; P:proteolysis; IEA.  
 KM Hydrolyase; Protease; Serine protease.  
 SQ SEQUENCE 412 AA; 44421 MW; CEBBF0782716EF8 CRC64;

Query Match 19.3%; Score 662; DB 2; Length 412;  
 Best Local Similarity 37.8%; Pred. No. 8.3e-28;  
 Matches 166; Conservative 72; Mismatches 155; Indels 46; Gaps 14;

QY 13 LVGLAGTALAAPVRYVAVNNAVQOKNYGLTPGLFKKYQAMNNOEVDVIMFGSYGRD 72  
 DB 3 LIPMFPSKIA--KEIVSNKIDKR--LLKALYLR-----SECVFVLY-SHLSY 48  
 QY 73 DRAVKTLRLMGAOVKSYKIPAVAVKIKARDLLIAGMIDTGFPGNTFVSGIKFOEDY 132  
 DB 49 DLKEKIEKGSGSIKPELIIKAMSVNLPCLDKNFA-----TLKIHIIAEDS 97  
 QY 133 KYQVVDATSVSQIADPTWNSLGYDSGVVAIVDGDIDANHPD---KGYIGYDAVN 189  
 DB 98 AVKLDQYATQEIARNA-NDLGYTGKVTIAFLDTGI-YPHPDFTKPKNRIVAFHDVN 155  
 QY 190 GRSTPYDDQGHGTTHVAGIYAGTG-SVNSQYIGVAPGAKIVGVKVLGADSGSVSTIAGV 248  
 DB 156 GKQGYDDNGHGTTHVAGIYAGTG-SVNSQYIGVAPGAKIVGVKVLGADSGSVSTIAGV 215  
 QY 249 DWVONKDKYKIRVINSLSGSSQSDG-TDSLQAVNNAMPAGIVVCVAAGNSGPNYTV 307  
 DB 216 QWILNKKKYNIRVNSLSIGETPSLPAPLDPVRGVDTLMRGLVTVAAAGNSGPNYTV 275  
 QY 308 GSPAAASRVITTYGAVD-----SNDNIASFSSRGPTADGRLEKPEVAVAGVDIIAPRASG 360  
 DB 276 TSPGTSKALIVGAVDDKRTSDISDETAQPSRG-SPLYKPDIVAPGVKIVSTASEN 333  
 QY 361 TSMG---TPINDYTKAGTSMATPHVSGVALIIQAPSWTPDKVKTALLETADIAPK 417  
 DB 334 VPFADDEVTLNKAYTATGTSMATPHMAGAAALLERKPNLTNVOIKILASTA--IKID 391  
 QY 418 ETADIAGAGRVNYKAIK 436  
 DB 392 DAGLMTQSGSMINTEALK 410

RESULT 13  
 O82B14 STRAW PRELIMINARY; PRT; 1208 AA.  
 AC O82B14;  
 DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.  
 DT 01-JUN-2003, sequence version 1.  
 DE 07-FEB-2006, entry version 19.  
 DE Putative protease.  
 GN OrderedLocNames=SAV5721;  
 OS Streptomyces avermitilis.  
 OC Bacteria; Actinobacteria; Actinomycetales; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomycetes.  
 NC NCBI\_Taxid=33903;  
 RN [1]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL B165;  
 RX MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;  
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,  
 RT "Genome sequence of an industrial microorganism Streptomyces  
 avermitilis: deducing the ability of producing secondary  
 metabolites.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL B165;  
 RX MEDLINE=22608306; PubMed=12692562; DOI=10.1038/nbt820;  
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
 RA Sakaki Y., Hattori M., Omura S.,  
 RT "Complete genome sequence and comparative analysis of the industrial  
 microorganism Streptomyces avermitilis.";  
 RL Nat. Biotechnol. 21:526-531(2003).  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs License  
 CC EMBL: BA000030; BAC73433.1; -; Genomic DNA.  
 DR HSSP; Q99405; IMP.  
 DR Biocyc: SAVE227882:SAV5721-MONOMER; -.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004289; F:subtilase activity; IEA.  
 DR GO; GO:000508; P:proteolysis; IEA.  
 DR InterPro; IPR003137; PA.  
 DR InterPro; IPR002029; Pept\_S8\_S53.  
 DR Pfam; PF02225; PA; 1.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 DR Complete proteome; Hydrolyase; Protease; Serine protease.  
 KW SEQUENCE 1208 AA; 125549 MW; E650B5E3AB5312B CRC64;  
 SQ  
 Query Match 19.1%; Score 657.5; DB 2; Length 1208;  
 Best Local Similarity 39.0%; Pred. No. 5.4e-27;  
 Matches 181; Conservative 73; Mismatches 169; Indels 41; Gaps 17;  
 29 VVRNNAVQKXNYGLTPGLFKVY---QRMNNQBVDTYIMGSGYDRDRAYKVLKMA 84  
 90 VVDEALPYLRAGTDLRLFDVSALIRQGLSDKGTGETPLIVG-YKGARAVTP--SGA 145  
 85 QVKYSYKILPAVAVKI-KARDLLILAGMIDTGYFGNTRVSGIKFIQEDYKVVQVDTATVS 143  
 146 TTRTSITSLGAAAVNDKGR-----TFKRSWTRGGIEHWLDAKREDAESNA 195  
 144 QIGADTVNSLIGDGGVVAIVDTGIDANHPDLKKGVIQWYDAVNGRSTPYDDQGHGH 203  
 196 QIGTRAAWDA-GLTGPDGVVAIVLDTGVTTHPDLAGRVSRKSFIDGEEVA-DRNGHGH 253  
 204 VAGIYAGTSV-NSQYIGVAPGAKLVGVKYLADGSGSVSTIIAGVDWVYQNKDKGIV 262  
 254 VTSYVGGSSAASDGTGRTGVAFGATLAVGVLSDDGSGSSOIIAGIEMARD---VRARI 310  
 263 IMLSGSSQSSDGTSLISOAVNN-AMWDAGIVVCAAGNSGPNRYTGVSPAAASKVITVA 321  
 311 VSMISLSTASDGTDPMAAVDITLSEETALFPVNAAGNTGAP8-SIGSGAADSALTVA 369  
 322 VDSNDNIASFSSRGPT-ADGRLLKEPVAVGVIIAPRASGTSMGTPINDYTKASGTSMA 380  
 370 VVSSDRAAFYTSAGPHGNALKPDIAAGGVDIRAAR-----SQLAPGTGYTSMGTSMA 425  
 381 TPHVSGVALLIIOAHSGWTPDKKTKLLIFTADIVAKETIADIVAGGRNVYKATIDY 440  
 426 TPHVGVALLLABQHPDWTGARKDLKMLMSTSEOL---DASVYQLQGRVSVDPAVG--- 478  
 441 AKLTFTGSVAADKGSATHTFDVSGATFVATLTYWDTGSSDIDLVL 484

DB 479 ARVATAGS-ADIGFRHPRHDPADRPVTKVT-VSNSSDITVELSL 520  
 RESULT 14  
 ID Q46C21\_METBA PRELIMINARY; PRT; 644 AA.  
 AC Q46C21;  
 DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.  
 DT 13-SEP-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 6.  
 DE Hypothetical protein.  
 GN OrderedLocustNames=Mbar\_A1626;  
 OS Methanococcus barkeri.  
 OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_Taxid=2208;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=Fusaro / DSM 804;  
 RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,  
 RA Hamon N., Israni S., Pittluck S., Goodwin L.A., Saunders E.H.,  
 RA Schmutz J., Larimer F., Land M., Anderson I., Richardson P.,  
 RT "Complete sequence of chromosome 1 of Methanosarcina barkeri str.  
 Fusaro.";  
 RT Submitted (Aug-2005) to the EMBL/GenBank/DBJ databases.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs License  
 CC EMBL: CP000099; AA270571.1; -; Genomic DNA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004289; F:subtilase activity; IEA.  
 DR GO; GO:000508; P:proteolysis; IEA.  
 DR InterPro; IPR002029; Pept\_S8\_S53.  
 DR InterPro; IPR00601; PKD.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR Pfam; PF00801; PKD; 2.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR SMART; SM00089; PKD; 2.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 KW Complete proteome; Hydrolyase; Hypothetical protein; Protease;  
 KW Serine protease.  
 KW SEQUENCE 644 AA; 67861 MW; D250F20D9E06629 CRC64;  
 SQ  
 Query Match 18.9%; Score 651; DB 2; Length 644;  
 Best Local Similarity 30.1%; Pred. No. 5.7e-27;  
 Matches 215; Conservative 95; Mismatches 234; Indels 170; Gaps 31;  
 4 LGAVALLVGLAGTALAAPVKKPVVRNNAVQKXNYGLTPGLFKVQRMNNQBVDTV 63  
 17 LCVICTSLILISLFFSLALAGDI-----NTSPINSDKISASVSKLREISGDEQIPVI 71  
 64 IMF-----GSY---GDRDAVYKLRIMQAVKYS-----YKILPAVAVKIKARL 105  
 72 ITLPNQKIAFNTAAKRYQIESQKNLIKPLENEKKNKVKIKIITINAAKVNPEVL 131  
 106 ILIAGMIDTGYFGNTRVSGIKFIQEDYKVVQVDTATVS----- 141  
 132 ASLAGRSVVS-----KIEIDVSVISVQSPSPFKKIAASTTSLTAAT 174  
 142 ---VSQIGADTVNSLIGDGGVVAIVDTGIDANHPDLK-----KVIGWYDA 187  
 175 NAMGVDKIDAPAVMOQ-GINGKGIYVAIVDTGIDATHPDLDDNPSTNDPKVVGWVDY 233  
 188 VNGRSTPYDDQGHGHVAGIYAGTSVNSQYIGVAPGAKLVGVKYLADGSGSVSTIIAG 247  
 234 INSQSAVDNDNGHGHVAGIYAGTSD-NGIQGVAPGKTLVAKPDSGDDYLSLCLIG 292  
 248 VDWVQNKDKYKGIYINLSLSSQSSDGTSL-SQAVNNAMDAIVVCAAGNSGPNRYT 306

Db 293 FEMAVANN---NARIISFSGSGSPH---DSLFTTMINKVVAAGVPIVIAAGNDGSGGT 344  
 QY 307 VCSPPAASAKVITGAVDSNDINIASFSRGP-TADGR--LKEPVAPGVDIAPRAGTSM 363  
 Db 345 ITPGPELNSLTGADDSALADFSRGPVITDDQYIKPDISARVSI-----P 395  
 QY 364 GRPINDYTKASGTSMATPHVSGVALLIQAHPSWTPDKVKYALIEFA-DIVAPKEIADI 422  
 Db 396 STVPGGQYAVGDGTSMAAPVSGTVALILEKPTMTPEAVKKELESTAVDLGSAGK--DN 453  
 QY 423 AYAGAVVNYKAL-----KYDDYAKLTFPGSVADKSGATTFPVSGATFTYATLTYMDTS 477  
 Db 454 DYSSGRDADAVAGKGPVAFYAKPT-SGKVPILTVAFTDT-----STGTPTRKRNFGD 507  
 QY 478 SDIDLVLTPDNGNEVDVSYAAVYGFKEVGYNPTAGTWTYK--VVSYKQ-----AANYQV 530  
 Db 508 GS-KSFLQNPK-----HKYS-----KAGTYTNLVKAKAGKQYTKYDIY 548  
 QY 531 DVVSDGSLSGSGGNPNPNPPTPTDTQTFTGSVN-DYMDTSDTFMNVNSGATKIT 569  
 Db 549 -VISKPTAARSA-----SPTSGKAPLTVAFDTKSSGSPNPAWKMS-----F 587  
 QY 590 GDITPDTSYNDLILYDPNGNL---VDRSTSSNSYEHVY---ANPAGTWT 636  
 Db 588 GDGTISREKNPTRQYLOEGKYKITLTVSNAAGSKTKKINYIKVTTTRBGIYS 641

## RESULT 15

OSJ315\_GEOKA PRELIMINARY; PRT; 442 AA.  
 ID OSJ315\_GEOKA  
 AC OSJ315;  
 DT 01-FEB-2005, integrated into UniProtKB/TrEMBL.  
 DT 01-FEB-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 10.  
 DE Intracellular alkaline serine protease.  
 GN OrderedLocustNames=GK0210;  
 OS Geobacillus kaustophilus.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.  
 OX NCBI\_TaxID=1462;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=HTA426;  
 RX PubMed=15576355; DOI=10.1093/nar/gkh970;  
 RA Takami H., Takaki Y., Chee G.-J., Nishi S., Shimamura S., Suzuki H., Matsui S., Uchiyama I.;  
 RA "thermoadaptation trait revealed by the genome sequence of thermophilic Geobacillus kaustophilus";  
 RT Nucleic Acids Res. 32:6292-6303(2004).  
 RL  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivative license  
 CC  
 DR EMBL, BA000043; BAD74495.1; -; Genomic DNA.  
 DR GO; GO:0008233; F:Dependase activity; IEA.  
 DR GO; GO:0042802; F:Protein self binding; IEA.  
 DR GO; GO:0004289; F:Subtilase activity; IEA.  
 DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.  
 DR GO; GO:0006508; P:Proteolysis; IEA.  
 DR InterPro; IPR000209; Pept\_S8\_S53.  
 DR InterPro; IPR010259; Prot\_inh\_S8A.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; UNKNOWN\_1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 KM Complete proteome; Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 442 AA; 48629 MW; 30F9F760ADD09D20 CRC64;

Query Match 18.9%; Score 649.5; DB 2; Length 442;  
 Best Local Similarity 40.6%; Pred. No. 4.3e-27;  
 Matches 152; Conservative 65; Mismatches 120; Indels 37; Gaps 11;

QY 85 QVKSYSKIIPAAVAKI--KARDILLIAGMIDTGYFGNTRVSGIKFIOEDYKQVDDATSV 142  
 Db 79 KYVHHRRHVPFYSARVTPALAEQLLEHPKVKYFENRT-----VKALLNNAVPS 127  
 QY 143 SQIGALTWNLSLGYDSGVVVAIVDNGIDANHPDLKGKYLGMWDVAVNGRSTPYDDQHGHT 202  
 Db 128 ANAKRVAV-NGTELSEKGVTLAVDTGI-YPHPDLEGRIAAFVDFVNGKRTTPYDDNGHGT 185  
 QY 203 HVAGIAGTGSV-NSOYIGVAPGAKLVGVKVLGADSGSVSTIIAGVDWVYVQNKDKY--- 258  
 Db 186 HCAGDAAGNRMSDGLIAGPAAVLANLIGVKVLDRSSGSGLETIMRGIEMWICIDVNEHPSK 245  
 QY 259 GIRVNLISLSSGSS--DGTSLSQAVNNAMDAGIVCYAAGNSGPNYTTVGSPPAAAK 315  
 Db 246 RIDIISLSIGGEPPQPEIENDDEPLVQVAAQAMEQGIIVCAAGNEGPYNGTISSPGISDR 305  
 QY 316 VITVGAVD-----SNDINIASFSRGPAPDRLEPVVAPGVDI---APRASGTSNG 364  
 Db 306 IITVGLADHDTRATTRADDVASFSSRGPTEYGVTPDLVPEVNILSLRAPSFLDKKN 365  
 QY 365 --TPINDYTKASGTSMATPHVSGVALLIQAHPSWTPDKVKYALIEFA-DIVAPKEIADI 422  
 Db 366 KQSRVGDHIIISNGTSMATPICAGIYALMLQAKPNATPDEIKRALDGDADLWKGRD--PV 423  
 QY 423 AYAGAVVNYKAIK 436  
 Db 424 VYAGAYVNGKRAIE 437

Search completed: January 6, 2007, 22:31:24  
 Job time: 194 secs

November 2005

Published\_Applications\_Nucleic Acid and Published\_Applications\_Amino Acid database searches now generate two sets of results each. The Published\_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published\_Applications\_New databases; older published applications make up the Published\_Applications\_Main databases.

Searches run against Nucleic Acid Published\_Applications produce two sets of results, with the extensions **.rnpbm** (Published\_Applications\_NA\_Main) and **.rnpbn** (Published\_Applications\_NA\_New).

Searches run against Amino Acid Published\_Applications produce two sets of results, with the extensions **.rapbm** (Published\_Applications\_AA\_Main) and **.rapbn** (Published\_Applications\_AA\_New).

**This Page Blank (uspto)**

**This Page Blank (uspto)**



GenCore version 5.1.9  
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2007, 22:33:57 ; Search time 31 Seconds  
(without alignments)  
2087.781 Million cell updates/sec

Title: US-10-800-684-1  
Perfect score: 3437  
Sequence: 1 MKRLGAVLVLTAVGLAGT.....YAVSTYGMADYQLKAVVYVG 659

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 393712 seqs, 98211237 residues  
Total number of hits satisfying chosen parameters: 393712

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Published Applications AA New:  
1: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/US09\_NEM\_PUB pep.\*  
2: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/US06\_NEM\_PUB pep.\*  
3: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/US07\_NEM\_PUB pep.\*  
4: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/US08\_NEM\_PUB pep.\*  
5: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/US10\_NEM\_PUB pep.\*  
6: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEM\_PUB pep.\*  
7: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEM\_PUB pep.\*  
8: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/US60\_NEM\_PUB pep.\*

\* Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID                  | Description       |
|------------|-------|-------------|--------|---------------------|-------------------|
| 1          | 2664  | 77.5        | 663    | US-10-526-324-251   | Sequence 251, App |
| 2          | 632.5 | 18.4        | 443    | US-10-530-643-8     | Sequence 8, Appl  |
| 3          | 615   | 17.9        | 1227   | US-10-530-643-6     | Sequence 6, Appl  |
| 4          | 576.5 | 16.8        | 552    | US-10-530-643-124   | Sequence 124, App |
| 5          | 559.5 | 16.3        | 379    | US-11-433-614-14    | Sequence 14, Appl |
| 6          | 546.5 | 15.9        | 1606   | US-10-530-643-114   | Sequence 114, App |
| 7          | 523.5 | 15.2        | 275    | US-10-567-073-3     | Sequence 3, Appl  |
| 8          | 523.5 | 15.2        | 275    | US-10-541-737-1     | Sequence 1, Appl  |
| 9          | 521.5 | 15.2        | 274    | US-10-532-605-2     | Sequence 2, Appl  |
| 10         | 521.5 | 15.2        | 274    | US-10-541-737-3     | Sequence 3, Appl  |
| 11         | 516.5 | 15.0        | 381    | US-11-452-695-2     | Sequence 2, Appl  |
| 12         | 513.5 | 14.9        | 381    | US-11-452-695-3     | Sequence 3, Appl  |
| 13         | 513.5 | 14.9        | 381    | US-11-452-695-4     | Sequence 4, Appl  |
| 14         | 508.5 | 14.8        | 524    | US-10-526-324-20    | Sequence 20, Appl |
| 15         | 507.5 | 14.6        | 381    | US-11-433-614-16    | Sequence 16, Appl |
| 16         | 500.5 | 14.6        | 275    | US-10-541-737-2     | Sequence 2, Appl  |
| 17         | 500   | 14.5        | 624    | US-10-530-643-46    | Sequence 46, Appl |
| 18         | 490.5 | 14.3        | 673    | US-10-530-643-56    | Sequence 56, Appl |
| 19         | 488   | 14.2        | 462    | US-10-449-902-40515 | Sequence 40515, A |
| 20         | 487.5 | 14.2        | 269    | US-10-541-737-4     | Sequence 4, Appl  |
| 21         | 486   | 14.1        | 274    | US-11-452-695-1     | Sequence 1, Appl  |
| 22         | 484   | 14.1        | 400    | US-10-530-643-60    | Sequence 60, Appl |
| 23         | 476.5 | 13.9        | 518    | US-10-530-643-122   | Sequence 122, App |
| 24         | 476   | 13.8        | 422    | US-10-526-324-979   | Sequence 979, App |
| 25         | 470.5 | 13.7        | 469    | US-10-449-902-37064 | Sequence 37064, A |

|    |       |      |     |   |                   |                   |
|----|-------|------|-----|---|-------------------|-------------------|
| 26 | 469   | 13.6 | 640 | 7 | US-11-318-576-4   | Sequence 4, Appl  |
| 27 | 468.5 | 13.6 | 523 | 6 | US-10-530-643-4   | Sequence 4, Appl  |
| 28 | 465   | 13.5 | 607 | 6 | US-10-530-643-62  | Sequence 62, Appl |
| 29 | 457.5 | 13.3 | 434 | 7 | US-11-318-576-10  | Sequence 10, Appl |
| 30 | 454.5 | 13.2 | 615 | 6 | US-10-530-643-88  | Sequence 88, Appl |
| 31 | 452.5 | 13.2 | 434 | 7 | US-11-318-576-2   | Sequence 2, Appl  |
| 32 | 452.5 | 13.2 | 434 | 7 | US-11-318-576-11  | Sequence 11, Appl |
| 33 | 450.5 | 13.1 | 434 | 7 | US-11-318-576-15  | Sequence 15, Appl |
| 34 | 443   | 12.9 | 433 | 7 | US-11-318-576-12  | Sequence 12, Appl |
| 35 | 443   | 12.9 | 433 | 7 | US-11-318-576-14  | Sequence 14, Appl |
| 36 | 438   | 12.7 | 269 | 7 | US-11-318-576-13  | Sequence 13, Appl |
| 37 | 433.5 | 12.6 | 269 | 6 | US-10-516-164A-1  | Sequence 1, Appl  |
| 38 | 432.5 | 12.6 | 433 | 7 | US-11-318-576-16  | Sequence 16, Appl |
| 39 | 432.5 | 12.6 | 611 | 6 | US-10-530-643-74  | Sequence 74, Appl |
| 40 | 429   | 12.5 | 434 | 6 | US-10-530-643-2   | Sequence 2, Appl  |
| 41 | 423   | 12.3 | 426 | 6 | US-10-530-643-42  | Sequence 42, Appl |
| 42 | 413   | 12.0 | 412 | 6 | US-10-530-643-128 | Sequence 128, App |
| 43 | 407   | 11.8 | 515 | 6 | US-10-530-643-50  | Sequence 50, Appl |
| 44 | 397.5 | 11.6 | 579 | 6 | US-10-530-643-66  | Sequence 66, Appl |
| 45 | 396.5 | 11.5 | 579 | 6 | US-10-530-643-70  | Sequence 70, Appl |

## ALIGNMENTS

RESULT 1  
US-10-526-324-251  
Sequence 251, Application US/10526324  
Publication No. US20060248617A1  
GENERAL INFORMATION:  
APPLICANT: Imanaka, Takayuki  
APPLICANT: Atom, Haryuki  
TITLE OF INVENTION: METHOD OF TARGETED GENE DISRUPTION, GENOME OF  
TITLE OF INVENTION: HYPERMOSTABLE BACTERIUM AND GENOME CHIP USING  
TITLE OF INVENTION: THE SAME  
FILE REFERENCE: 490051.401USPC  
CURRENT FILING DATE: 2005-02-28  
PRIOR APPLICATION NUMBER: US/10/526,324  
PRIOR FILING DATE: 2003-08-29  
PRIOR APPLICATION NUMBER: PCT/IB2003/003597  
PRIOR FILING DATE: 2002-08-29  
PRIOR APPLICATION NUMBER: JP 2002-319011  
PRIOR FILING DATE: 2002-08-30  
NUMBER OF SEQ ID NOS: 2167  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 251  
LENGTH: 663  
TYPE: PRT  
ORGANISM: Thermococcus kodakarensis KOD1  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (414562)..(414542)  
OTHER INFORMATION: n is a or c or g or t.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (786907)..(786907)  
OTHER INFORMATION: n is a or c or g or t.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (786907)..(786907)  
OTHER INFORMATION: n is a or c or g or t.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (786907)..(786907)  
OTHER INFORMATION: n is a or c or g or t.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (839139)..(839139)  
OTHER INFORMATION: n is a or c or g or t.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1128488)..(1128488)  
OTHER INFORMATION: n is a or c or g or t.  
FEATURE:

```

NAME/KEY: misc feature
LOCATION: (1128499)..(1128499)
OTHER INFORMATION: n is a or c or g or t.
FEATURE:
NAME/KEY: misc feature
LOCATION: (1128505)..(1128506)
OTHER INFORMATION: n is a or c or g or t.
FEATURE:
NAME/KEY: misc feature
LOCATION: (1128517)..(1128518)
OTHER INFORMATION: n is a or c or g or t.
FEATURE:
NAME/KEY: misc feature
LOCATION: (1128539)..(1128540)
OTHER INFORMATION: n is a or c or g or t.
FEATURE:
NAME/KEY: misc feature
LOCATION: (1561400)..(1561400)
OTHER INFORMATION: n is a or c or g or t.
FEATURE:
NAME/KEY: misc feature
LOCATION: (1561477)..(1561477)
OTHER INFORMATION: n is a or c or g or t.
FEATURE:
NAME/KEY: misc feature
LOCATION: (1561545)..(1561545)
OTHER INFORMATION: n is a or c or g or t.
FEATURE:
NAME/KEY: misc feature
LOCATION: (1767941)..(1767941)
OTHER INFORMATION: n is a or c or g or t.
US-10-526-324-251

```

```

Query Match 77.5%; Score 2664; DB 6; Length 663;
Best Local Similarity 75.9%; Pred. No. 1,3e-181;
Matches 505; Conservative 64; Mismatches 88; Indels 8; Gaps 4;

```

```

QY 1 MKRGAVVLAALVGLAGRLAALPVKPVVRRNNAVQCKNTGLTTPGLFKYVQRMNNQEV 60
DB 1 MKRGAVVLAALVGLAGRLAALPVKPVVRRNNAVQCKNTGLTTPGLFKYVQRMNNQEV 59
QY 61 DTVMFSGYGRDRRAVAVKRLMGAQVYKXKIPAVAVKIKAPDLIIAAMIDTGYGNT 120
DB 60 STIMFNOADKEKAVILDFLAKIKYNTIIPALAVKIKVLDLIIAGIMDTGYFGNA 119
QY 121 RVSGIKRQEDYKVVQVDAT----SVSOIGADTVNNSLGYDSGVVAIVDTGIDANHP 175
DB 120 QLSGVCFIQEDYVVKVAVETEGDESAQVMAITMMN-LGYDSGITIGIIDGIDASHP 178
QY 176 DLGKVIQWDAVNGRSTPYDDOGHGHVAGIYAGTGSV-NSQYIGVAPAKIVGYKVLG 234
DB 179 DLGKVIQWDAVNGRSTPYDDOGHGHVAGIYAGTGSV-NSQYIGVAPAKIVGYKVLG 234
QY 235 ADGSGSVSTIIAGVWVONKDKYKIRVIMLSGSSOSSDGTDSLQAVANNADAGIYVC 244
DB 239 GQSGSGISDIINGVDAVONKDKKIGIVIMLSGSSOSSDGTDSLQAVANNADAGIYVC 298
QY 295 VAAAGSGPNTYVGSPPAASKVITVGAVDSNDNIASFSSRGPTADGRLKPEVVAPEGVDII 354
DB 299 VAAAGSGPNTYVGSPPAASKVITVGAVDSNDNIASFSSRGPTADGRLKPEVVAPEGVDII 358
QY 355 APAASGSMGTPIINDYVTKASGSMATPHYSYGALILQAHPSWTDPKVTALILEADIV 414
DB 359 APAASGSMGTPIINDYVTKASGSMATPHYSYGALILQAHPSWTDPKVTALILEADIV 418
QY 415 APKEIADIAAGRVVNYKAIKYDYAKLFTGSAVDKGSATHTFVSGATFTATLTYMD 474
DB 419 KPDEIADIAAGRVVNYKAIKYDYAKLFTGSAVDKGSATHTFVSGATFTATLTYMD 478
QY 475 TGSDDIDLTYDNGNEVDYSYTAAYGFEKVGYNPTAGTWTVKVSYKGAANYQVDVVS 534
DB 479 NSGSDIDLTYDNGNEVDYSYTAAYGFEKVGYNPTAGTWTVKVSYKGAANYQVDVVS 538

```

```

QY 535 DGSLSOSGGGNPNPNPNTPTTDPOTFNGSYNDYDSDTFETMNVNSGATKITGDTLP 594
DB 539 DGSLSOSGGSGSEBSPSEPTVDEKTFGVHDYDSDTFTMTVNSGATKITGDTLP 598
QY 595 DTSYNDIDLTYDNGNLVDRSTSSNSYEHVEYANPAPGTMTFLVYASTYGMADYOLKA 654
DB 599 DTSYNDIDLTYDNGNLVDRSTSSNSYEHVEYANPAPGTMTFLVYASTYGMADYOLKA 658
QY 655 VVYVG 659
DB 659 KVVYVG 663

```

## RESULT 2

```

US-10-530-643-8
Sequence 8, Application US/10530643
Publication No. US2006025995A1
GENERAL INFORMATION:
APPLICANT: CAYOUEITE, Michelle
APPLICANT: HANSEN, Connie Jo
APPLICANT: MCCLURE, Amy
APPLICANT: SUN, May
APPLICANT: GRAMATIKOVA, Svetlana
APPLICANT: DYCAICO, Mark
APPLICANT: BARTON, Nelson R.
APPLICANT: STEGE, Justin T.
APPLICANT: ABOUSHADI, Nabla M.
TITLE OF INVENTION: PROTEASES, NUCLEIC ACIDS ENCODING THEM
TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
FILE REFERENCE: 564462004100
CURRENT APPLICATION NUMBER: US/10/530,643
PRIOR FILING DATE: 2006-04-07
PRIOR APPLICATION NUMBER: PCT/US03/32819
PRIOR FILING DATE: 2003-10-10
PRIOR APPLICATION NUMBER: 60/471,423
PRIOR FILING DATE: 2003-05-16
PRIOR APPLICATION NUMBER: 60/418,467
PRIOR FILING DATE: 2002-10-10
NUMBER OF SEQ ID NOS: 255
SOFTWARE: FaastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 443
TYPE: PRT
ORGANISM: Bacteria
US-10-530-643-8

```

```

Query Match 18.4%; Score 632.5; DB 6; Length 443;
Best Local Similarity 36.8%; Pred. No. 3.6e-37;
Matches 153; Conservative 79; Mismatches 143; Indels 41; Gaps 12;

```

```

QY 47 LFKYVQRMNNN-QEYDVTVMFSGYGRDRRAVAVK---LRLMGAQVYKXKIPAVAVKIK 101
DB 39 LHKMERFRPKNNKKLSVILFEFSNCYQSGCLEVNOAFTRKNCCKIRNEFSIISCSADIT 98
QY 102 ARDLLIIAGMIDTYFGNTRVSGIKFQEDYKVVQ-VDDATSYSQIGADTVNNSLGYDSG 160
DB 99 PSLVEEVL-----TNCNHIKVVYLNEHVRALLDTAVISANAKNIVRNNTLTGKG 148
QY 161 VVVAIVDVGIDANHPDLGKVIQWDAVNGRSTPYDDOGHGHVAGIYAGTGSVNS-QYI 219
DB 149 ITIAVYIDGI-FPHDTLSGRILDFPDFTNDRKREYDNDGHHGHCGLAGDSASGLYM 207
QY 220 GVAAPAKLVGVKVGADSGSVSTIIAGVWV---ONKDYKIRVIMLSGSS--SQS 272
DB 208 GVAPEANVYGVVKNKVSGLSETVMGVDMCKIKNDQNPQ-KINIIMSLGAPAQRYE 266
QY 273 SGTGTSLSQAVNNANDAGIYVCVAAAGSGPNTYVGSPPAASKVITVGAVD-----S 324
DB 267 NENDDPMTVMYKAWENGIYVCVAAAGNEGPASTIASGVEOVITVGALDXTADTRS 326
QY 325 NNINASFSSRGPTADGRLKPEVVAPEGVDIIAPRASGTM-----GTPINDYVTKASGTM 379
DB 327 DDEVASFSSRGPTIYKVKFDIAPGVDIISLRSPNSYLDKRFQKSNRVGSDYFSLSGTSM 386

```

QY 380 APHVSQVGLITLOAHPSWTPDKVTALJETADIVAPKEIADIAVGARVNYKAI 435  
DB 387 AFDICAGIALILIQHNPNATPOEVKOLLRKTDLTNRD--PNYVAGIYIMNSV 440

## RESULT 3

US-10-530-643-6  
; Sequence 6, Application US/10530643  
; Publication No. US2006025995A1  
; GENERAL INFORMATION:  
; APPLICANT: CAYOJETTE, Michelle  
; APPLICANT: HANSEN, Connie Jo  
; APPLICANT: MCCURE, Amy  
; APPLICANT: SUN, May  
; APPLICANT: GRAMATIKOVA, Svetlana  
; APPLICANT: DYCAICO, Mark  
; APPLICANT: BARTON, Nelson R.  
; APPLICANT: STEGE, Justin T.  
; APPLICANT: ABOUSHADI, Nahla M.  
; TITLE OF INVENTION: PROTEASES, NUCLEIC ACIDS ENCODING THEM  
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM  
; FILE REFERENCE: 564462004100  
; CURRENT FILING DATE: 2006-04-07  
; PRIOR FILING DATE: 2003-10-10  
; PRIOR APPLICATION NUMBER: PCT/US03/32819  
; PRIOR FILING DATE: 2003-10-10  
; PRIOR APPLICATION NUMBER: 60/471,423  
; PRIOR FILING DATE: 2003-05-16  
; PRIOR APPLICATION NUMBER: 60/418,467  
; PRIOR FILING DATE: 2002-10-10  
; NUMBER OF SEQ ID NOS: 255  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 1227  
; TYPE: PRT  
; ORGANISM: Bacteria  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)...(32)  
US-10-530-643-6

Query Match 17.9%; Score 615; DB 6; Length 1227;

Best Local Similarity 43.3%; Pred. No. 2.6e-35; Matches 155; Conservative 55; Mismatches 122; Indels 26; Gaps 13;

QY 131 DYKVVQVDDATSVSOIGADITVWNSLGYDGSQVVAIVDTGIDANHPDLKGKVIQWYAVNG 190  
DB 192 DGRVKAAMESNMQITTPKAMEA-GILTKGVKVAIVDTGIDLHPDLAGRVTESKSFVAG 250  
QY 191 RSTPYDDQGHGHTVAGIVAGTGS-VNSQYIGVAPAKLVGVKVLGADGSGSVSTIIAGVD 249  
DB 251 QEVA-DRNGHGTHTVASTVGSAGSDGKEKGVAPGATLAVGKVLSDGSGSSEIITAGME 309  
QY 250 WVVQNDKXGIRVINSLSGSSSDGTDSLSCAVNN-AMDAQIVVCVAAANGS-PNTYTY 307  
DB 310 WAAKDD---AKVSNLSGSRPSDGTDPVALVNLMTETGALFVYIAGNSGYPG-SI 364  
QY 308 GSPAAASKYITVAVNSNDNIASFSSRGPT-ADGRLEKEPVAVGVDIAPRASGTSMTGP 366  
DB 365 GSPGADSAALITGIVADSBAAVFTSGRPYGDQALKPDLISAPGVIIAAR---SGLLP 420  
QY 367 INDYTKASGTSNATPHVSGVGLIILOAHPSWTPDKVTALJETADIVAPKEIADIAVGA 426  
DB 421 GSGLYTSMGTSNATPHVAGVALLAERHPDWTGADLKALMSSSKYL---DASSYALGS 477  
QY 427 GRNVVVKAIKYDDYAKLTFGTGVAADKGSATHTFDVSGATVTLNLTWDGSSDIDLYL 484  
DB 478 GRVDAALIA---ANVTATGS-ADIGFVAMPYASSKPVTKVTV---TMSDAPVEL 527

RESULT 4  
US-10-530-643-124

; Sequence 124, Application US/10530643  
; Publication No. US2006025995A1  
; GENERAL INFORMATION:  
; APPLICANT: CAYOJETTE, Michelle  
; APPLICANT: HANSEN, Connie Jo  
; APPLICANT: MCCURE, Amy  
; APPLICANT: SUN, May  
; APPLICANT: GRAMATIKOVA, Svetlana  
; APPLICANT: DYCAICO, Mark  
; APPLICANT: BARTON, Nelson R.  
; APPLICANT: STEGE, Justin T.  
; APPLICANT: ABOUSHADI, Nahla M.  
; TITLE OF INVENTION: PROTEASES, NUCLEIC ACIDS ENCODING THEM  
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM  
; FILE REFERENCE: 564462004100  
; CURRENT FILING DATE: 2006-04-07  
; PRIOR FILING DATE: 2003-10-10  
; PRIOR APPLICATION NUMBER: PCT/US03/32819  
; PRIOR FILING DATE: 2003-10-10  
; PRIOR APPLICATION NUMBER: 60/471,423  
; PRIOR FILING DATE: 2003-05-16  
; PRIOR APPLICATION NUMBER: 60/418,467  
; PRIOR FILING DATE: 2002-10-10  
; NUMBER OF SEQ ID NOS: 255  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 124  
; LENGTH: 552  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample.  
; NAME/KEY: SIGNAL  
; LOCATION: (1)...(56)  
US-10-530-643-124

Query Match 16.8%; Score 576.5; DB 6; Length 552;

Best Local Similarity 43.7%; Pred. No. 4.8e-33; Matches 141; Conservative 37; Mismatches 84; Indels 61; Gaps 9;

QY 158 GSGVVAIVDTGIDANHPDLKG-----KYIGWYDAVNGSTPYDDQGHGHTVVA 205  
DB 245 GSGVTALIIDTGVDTYHPLDGGCTTQEFPGACEKVIQYDFINDDAMDNDGHTVVA 304  
QY 206 GIVAGTGSVNSQYIGVAPAKLVGVKVLGADGSGSVSTIIAGVDWVYQNDKXGIRVIN 265  
DB 305 GIAAGNGILK---GVAAPAKTILAYKVLGAGGVGTWEGIIAGIEQAVID---GADILSL 356  
QY 266 SIGSSQSDGTSL-SQAVNNAMDAGIVVCVAAANGSPNTYTVGSPAAASKYITVGAVD 324  
DB 357 SIGCHSSCNPDIDIAQAVDNNAVLAGKVYVVAANGSPSSRTIGSGGTARKAITVGSITK 416  
QY 325 NNINIAFSSRGPTA-----DGRLEKEPVAVGVDIAPRASGTSMTGP----- 366  
DB 417 SIIISFSSRGFPVMMDBAGIEQALMKPDVLPAG-----GTDGSEFNCPEMFDN 467  
QY 367 -----INDYTKASGTSNATPHVSGVGLIILOAHPSWTPDKVTALJETADIVA--PKG 418  
DB 468 RICAAMLINEYLAISGTSNATPLVSGAIALKQKHDPWTPEELKGAIVKGTAINLGYDPNE 527  
QY 419 IADIAVGAGRVVYKAIKYDDYA 441  
DB 528 -----QGAQRINVRWITGLEBRA 545

RESULT 5  
US-11-433-614-14  
; Sequence 14, Application US/11433614  
; Publication No. US20060205019A1  
; GENERAL INFORMATION:  
; APPLICANT: Athena Biotechnologies, Inc.  
; APPLICANT: Menzel, Rolf  
; TITLE OF INVENTION: Methods and Compositions for Directed Gene Assembly

```

; FILE REFERENCE: 46675-5004-01-US
; CURRENT APPLICATION NUMBER: US/11/433,614
; CURRENT FILING DATE: 2006-05-12
; PRIOR APPLICATION NUMBER: US 09/920,118
; PRIOR FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent version 3.3
; SEQ ID NO 14
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-11-433-614-14

```

```

Query Match      16.3% Score 559.5; DB 7; Length 379;
Best Local Similarity 38.2%; Pred. No. 4,66-32;
Matches 147; Conservative 54; Mismatches 135; Indels 49; Gaps 13;

```

```

QY 57 NOEVDVTIMFGSYGDRDAVK--VLRLMGAQVKTYSKTIIPAVAKIKARDLLIAGMIDT 114
DB 35 NVEKDVIYGFKS-GVKTKASVKKDIKSGGKVDKQFRILINAKAKLDEALEEYKNDPDV 93
QY 115 GFQNTRVSGIKFIOEDYKQVDDAT--SVSOIGADTVWNSLGYDGSVVVAIVDTGID 171
DB 94 AV-----VEBHVAAHALAQTVPYGIPLIKADKV-QAOSYKGAANKVAVALDTGIQ 141
QY 172 ANHPDLKGVIGWYDAVNGRSTPYDDQGHGTHVAGIAGTGSVNSOYIGVAPGAKLVGVK 231
DB 142 ASHDDL--NVVGGASFYVAGEAYNTDGNHGHVAGTVAALDNTTG-VLGVAPVNSLYAVK 198
QY 232 VLGADGSGSVSTIIAGVDMVWQNKDKGIRVINSLSGSSQSDGTDLSQAVNNADAGT 291
DB 199 VINSGSGSYSGISGISEMATTN---GMDVINMSLD--GPSSTIMKQAVNNAAYARGV 251
QY 292 VVCAAGNSGP--NTYTVGSPAAASKYITVGAVDNSNINASFSSRGPTADGRLEKPEVAP 349
DB 252 VVVAAGNSGSGSNTNTIGYPAKYDSIYAVGAVDPNRRASFSSVG-----ALEVMAP 305
QY 350 GVDIARASGTSNGTINDYTTASGTSNATPHVSGVALIIOAHPSWTPDKVKALIE 409
DB 306 GAGVYSTYPTST-----YATLNGTSMASPHVAGAAALILSKHPNLSASQVRRLSS 356
QY 410 TADIVAPKEIADIYAGGAVNVVYKA 434
DB 357 TATYLG---SPTYGKGLINVEAA 377

```

```

RESULT 6
US-10-530-643-114
; Sequence 114, Application US/10530643
; Publication No. US2006025995A1
; GENERAL INFORMATION:
; APPLICANT: CAYOUEITE, Michelle
; APPLICANT: HANSEN, Connie Jo
; APPLICANT: MCCLURE, Amy
; APPLICANT: SUN, May
; APPLICANT: GRAMATIKOVA, Svetlana
; APPLICANT: DYCAICO, Mark
; APPLICANT: BARTON, Nelson R.
; APPLICANT: STEEG, Justin T.
; APPLICANT: ABOUSHADI, Nahla M.
; TITLE OF INVENTION: PROTEASIS, NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 564462004100
; CURRENT APPLICATION NUMBER: US/10/530,643
; CURRENT FILING DATE: 2006-04-07
; PRIOR APPLICATION NUMBER: PCT/US03/32819
; PRIOR FILING DATE: 2003-10-10
; PRIOR APPLICATION NUMBER: 60/471,423
; PRIOR FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: 60/418,467
; PRIOR FILING DATE: 2002-10-10
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 114
; LENGTH: 1606
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample.
; NAME/KEY: SIGNAL
; LOCATION: (1)...(32)
US-10-530-643-114

```

```

Query Match      15.9% Score 546.5; DB 6; Length 1606;
Best Local Similarity 23.4%; Pred. No. 2,96-30;
Matches 240; Conservative 110; Mismatches 222; Indels 455; Gaps 41;

```

```

QY 7 VIALIVLVLGIALTAAPYKPVVNN--AVQK---NYGLTPG--LFFKVGQM 54
DB 16 LINVLSLVLPFNTANAAALLKQPFKTNPDASALQOKAIAQOTSLLGEDARLHKDQL 75
QY 55 NMNQEVDVTIMFGSYGDRDAVK--VLRLMGAQVK----- 87
DB 76 SSGQEVPIIHS--EKAVLEQGIHKLNGKKMSQSEITLKKGIQAQONTAKEMT 130
QY 88 -----YSY-KIIPAVAKIKARDLLIAGMIDTYGFNTRVSGIKFIOEDYKV--- 134
DB 131 VKKVKFPDYVSDTVLNGLSGTVOANDLKL-----LITSGVKVPEPTTVVAQ 179
QY 135 ----QVD---DA--TSVSOIGADTVWNSLGYDGSVVVAIVDTGIDANPDLK-----G 179
DB 180 EQLQKVPYVDKMDTSLPFLGINKLMBE-GIEGQGVKVAIVDTGIDASHPDLKAVYKGG 238
QY 180 K-----VIGWY-----DAVNGRSTPYD-----DQ-----HGTVAIGVAGTGSV 214
DB 239 KMFVHVDSFYARPRADDGSETSPIDRANKPEVNERSSYTTHTGHVAGTIVAQGN 298
QY 215 NSQYIGVAPGAKLVGVKVLGADGSGSVSTIIAGVDMVWQNKDKGIRVINSLSGSSQSD 274
DB 299 EFGIKGIAPKVDLYGVRVLGAVGSGATGIIKAIDTAVIEK---IDVINLSGGAGNSE 354
QY 275 GTDLSLSQAVNNAMDGIYVCVAAAGNSGRTTYVSGPAAASKYITVG----- 320
DB 355 -TDGASFALNNMLAGTISVATGNSGPRGTGTPATSRIGIAVGNTPNETHQAVNN 413
QY 321 ----- 320
DB 414 VKAGEFTYSKVNDLMATTFGKQVETQLAGYDLIAPGAGAKYADLDVTGKALVARS 473
QY 321 ---AVD----- 323
DB 474 EIAFYDKVAAAKKNGAVAIIVNFPAGTNAAPNKSDVFLGDAFEFIPFDMGVTDOEALRA 533
QY 324 -----SNDNIASFSSRGP-TADGSLKPEVYAPGVDTIA--PRASGT 361
DB 534 AIAAAPSTISPDQSSIKTTGDVDNDSSRGSPFNPFDIKPVSAFGTNISSTIPMYND 593
QY 362 SMGTINDYTTASGTSNATPHVSGVALIIOAHPSWTPDKVKALITADIVAPKEIAD 421
DB 594 PPDADYSRAYSARKTQSMATHIAGIALVQGANPTMNAFPVKYALSTATAIVLDTKKYD 653
QY 422 IAYGAGRVNVTYKAYD----- 438
DB 654 PAQAGRYDAYKARADVLAYAIDTASNDGTEVENLKGTVTFGPQKLDKNISYTKINVK 713
QY 439 -----DYA-----KLTFTGSVAD----- 451
DB 714 DKSAKGDIYTVSDVTYGFQDAKATVDQSEFTLNGEQLLVTLTASAKETAGDEILGIY 773
QY 452 --KG-----SATHTFDVSGATFVT-----ATLYW-----DTGSSD 479
DB 774 HIKGNGKDLSPFPAADFGAAVAVKOMEIKTKDLSFNGSGVNDADALYFTITDVGFN 833
QY 480 IDLY-LIDPNENEDVYSTAYYGEKVGYYNP---TAGTWTV---KVVSYKGAANYOV 530

```

```

Db      834 IBLMDIMNPTGKSYDGY-----IGYHASDITLTAAGSYCLPLIKYTOGWTAS--- 882
QY      531 DAVSDGSL-----SOSGGNPNPNPN-PMPTPTTDQ--FTTGSVND-----YMD- 572
Db      883 EVIIPDELTYIDPSAETKSNPRTIADYVGPVVVKSTAGTIEGAADGKVTGKTIDKRYDY 942
QY      573 -----TSDFTMNVN-----SGATKITGD-----LTFDTSYNDLDLY 604
Db      943 QKELVKYGWGYDLNLTSLATYEVTANDAVVDGAVVLADGTFAPFLPTFDKTKNNVTYK 1002
QY      605 LYDPNGN 611
Db      1003 YADAAGN 1009

```

```

RESULT 7
US-10-567-073-3
; Sequence 3, Application US/10567073
; Publication No. US20060134740A1
; GENERAL INFORMATION:
; APPLICANT: Bryan, Philip N.
; TITLE OF INVENTION: Engineered Proteases for Affinity Purification and Processing of
; TITLE OF INVENTION: Fusion Proteins
; FILE REFERENCE: 4115-181
; CURRENT APPLICATION NUMBER: US/10/567,073
; CURRENT FILING DATE: 2006-02-03
; PRIOR APPLICATION NUMBER: US 60/493,032
; PRIOR FILING DATE: 2003-08-06
; PRIOR APPLICATION NUMBER: PCT/US04/021049
; PRIOR FILING DATE: 2004-06-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Bacillus amyloliquefaciens
US-10-567-073-3

```

```

Query Match      15.2%; Score 523.5; DB 6; Length 275;
Best Local Similarity 44.4%; Pred. No. 1.1e-29;
Matches 132; Conservative 40; Mismatches 90; Indels 35; Gaps 11;

```

```

QY      142 VSOIGADTVWNSLGYDGSVVVAIVDTGIDANHPDLKGVIQWYDAVNGRSTPY--DDQGH 200
Db      8 VSOIKAPAL-HSOGYTGSNVKAVIDSGIDSHPLD--KVAGASVNPSETNPFQDNNSH 64
QY      201 GTHVAGIVAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGSVSTIIAGVDVWVQNKDKYGI 260
Db      65 GTHVAGTVAALNN-SIGVLGVAPASALYAVKVLGADGSGQYSWIINGIEMALAN---NM 119
QY      261 RVINISLSSGSSDGTDSLQAVNNAMDAIYVCVAAGNSGP--NTYTVGSPAAASKVIT 318
Db      120 DVINNLSLG---GPSGSAALKAADVDAVAGVVVAAGNEGTSSTVGYGPKYPSVLA 176
QY      319 VCAVNSNDNIASFSSRGPTADRLKPEVYAPGVDIIPASGTSMTGPIINDYYTKASGTS 378
Db      177 VCAVNSNORASFSSVGPBLD-----VMAPGVSI-----QSTLPGNKTYAGVNGTS 221
QY      379 MATPHVSGVALILQAHPSWTPDKVTALIFETADIAPKEIAD-IAYGAGRNVVYKA 434
Db      222 MASPHVAGAAALILSKHPMTWNTQVRSLENTT-----TKLGDSPFYGGKLINVQA 273

```

```

RESULT 8
US-10-541-737-1
; Sequence 1, Application US/10541737
; Publication No. US20060252155A1
; GENERAL INFORMATION:
; APPLICANT: Leeifang, Chris
; APPLICANT: Van der Kleij, Wilhelmus A.H.
; TITLE OF INVENTION: Methods for Site-Directed Mutagenesis
; TITLE OF INVENTION: and Targeted Randomization
; FILE REFERENCE: GC787-2

```

```

; CURRENT APPLICATION NUMBER: US/10/541,737
; CURRENT FILING DATE: 2005-07-08
; PRIOR APPLICATION NUMBER: PCT/US04/01334
; PRIOR FILING DATE: 2004-01-16
; PRIOR APPLICATION NUMBER: US 60/440,792
; PRIOR FILING DATE: 2003-01-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Bacillus amyloliquefaciens
US-10-541-737-1

```

```

Query Match      15.2%; Score 523.5; DB 6; Length 275;
Best Local Similarity 44.4%; Pred. No. 1.1e-29;
Matches 132; Conservative 40; Mismatches 90; Indels 35; Gaps 11;

```

```

QY      142 VSOIGADTVWNSLGYDGSVVVAIVDTGIDANHPDLKGVIQWYDAVNGRSTPY--DDQGH 200
Db      8 VSOIKAPAL-HSOGYTGSNVKAVIDSGIDSHPLD--KVAGASVNPSETNPFQDNNSH 64
QY      201 GTHVAGIVAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGSVSTIIAGVDVWVQNKDKYGI 260
Db      65 GTHVAGTVAALNN-SIGVLGVAPASALYAVKVLGADGSGQYSWIINGIEMALAN---NM 119
QY      261 RVINISLSSGSSDGTDSLQAVNNAMDAIYVCVAAGNSGP--NTYTVGSPAAASKVIT 318
Db      120 DVINNLSLG---GPSGSAALKAADVDAVAGVVVAAGNEGTSSTVGYGPKYPSVLA 176
QY      319 VCAVNSNDNIASFSSRGPTADRLKPEVYAPGVDIIPASGTSMTGPIINDYYTKASGTS 378
Db      177 VCAVNSNORASFSSVGPBLD-----VMAPGVSI-----QSTLPGNKTYAGVNGTS 221
QY      379 MATPHVSGVALILQAHPSWTPDKVTALIFETADIAPKEIAD-IAYGAGRNVVYKA 434
Db      222 MASPHVAGAAALILSKHPMTWNTQVRSLENTT-----TKLGDSPFYGGKLINVQA 273

```

```

Query Match      15.2%; Score 521.5; DB 6; Length 274;
Best Local Similarity 41.9%; Pred. No. 1.5e-29;
Matches 131; Conservative 39; Mismatches 94; Indels 49; Gaps 10;

```

```

QY      124 GIKFTQEDYKQVDDATSVSQIGADTVWNSLGYDGSVVVAIVDTGIDANHPDLKGKVI 183
Db      7 GIPLIKAD-KYQAQ-----CYKXANVYVGIIDTGIASSHDL--KVVG 46
QY      184 WYDAVNGRSTPYDDQGHGTHVAGIYAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGSVST 243
Db      47 GASPFSGESYNDGNGHGHVAGTVAALDNTTG-VLGVAPNVSLAIKVLNSSGSGTYSA 105
QY      244 IIAQVDWVWQNKDKYIRVINISLSSGSSDGTDSLQAVNNAMDAIYVCVAAGNSGP- 302

```

```
Db 106 IYSGIMATON-----GIDVTNMSIG---GPSGSTALKQADKAYASGIVVAAAGNSGSS 158
QY 303 -NNTYTGSPAAASKVITVGVDSNDNIASFSSRGPTADGRLKPEVAPGVDDIAPRASGT 361
Db 159 GSONTTGYPKYDSVLAAGVADSNKNRASFSSVGS-----ELEVMAFGVSSTPSNT 212
QY 362 SMGTPINDYTYKASGTSMATPHYSGVGALLIOAHPSWTDPKVTALLETADIAPKEIAD 421
Db 213 -----YTSLNGTSMASPHVAGAAALILSKYPTLSAQVNRRLSTATNIGD-----S 259
QY 422 IAYGAGRVTNYKA 434
Db 260 FYYGKGLINVEAA 272
```

RESULT 10  
US-10-541-737-3

```
/ Sequence 3, Application US/10541737
/ Publication No. US20060252155A1
/ GENERAL INFORMATION:
/ APPLICANT: Leeeflang, Chris
/ APPLICANT: Van der Kleij, Wilhelmus A.H.
/ TITLE OF INVENTION: Method for Site-Directed Mutagenesis
/ TITLE OF INVENTION: and Targeted Randomization
/ FILE REFERENCE: GC787-2
/ CURRENT APPLICATION NUMBER: US/10/541,737
/ PRIOR FILING DATE: 2005-07-08
/ PRIOR APPLICATION NUMBER: PCT/US04/01334
/ PRIOR FILING DATE: 2004-01-16
/ PRIOR APPLICATION NUMBER: US 60/440,792
/ PRIOR FILING DATE: 2003-01-16
/ NUMBER OF SEQ ID NOS: 6
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 3
/ LENGTH: 274
/ TYPE: PRT
/ ORGANISM: Bacillus licheniformis
US-10-541-737-3
```

```
Query Match 15.2%; Score 521.5; DB 6; Length 274;
Best Local Similarity 42.2%; Pred. No. 1.5e-29;
Matches 132; Conservative 39; Mismatches 93; Indels 49; Gaps 11;

QY 124 GIKPIODYKYQYVDATSVSOGIADVTWNSLGYDGSVNSQIYGVAPGAKLVGVKVLGADGSGSVT 183
Db 7 GIPILIKAD-KVQAO-----GFGANVYKVAVDITGLOASHPL--NVVG 46
QY 184 WYDAVNGRSTPYDQGHGTHVAGIYAGTGSVNSQIYGVAPGAKLVGVKVLGADGSGSVT 243
Db 47 GASFVAGEAVNTDNGHGTAVAGTVALDNTTG-VLGAVPSVSLYAVKVLNSGSGSVSG 105
QY 244 ILAGVDMVQNKXGIRVNLISGSSQSDGTDLSQAVNNMADAIYVCVAAAGNSG-- 301
Db 106 IYSGIMATON-----GMDVTNMSIG---GASGTAMQAVDANVAVVAAAGNSGNS 158
QY 302 PNTYVSGPAAASKVITVGVADSNDIASFSSRGPTADGRLKPEVAPGVDDIAPRASGT 361
Db 159 GSTTITIGYPKYDSVLAAGVADSNKNRASFSSVGS-----AELEVMAFG-----AGV 204
QY 362 SMGTPINDYTYKASGTSMATPHYSGVGALLIOAHPSWTDPKVTALLETADIAPKEIAD 421
Db 205 YSTYPTVTYAT-INGTSMASPHVAGAAALILSKHPNLSAQVNRRLSTATNIGD-----SS 259
QY 422 IAYGAGRVTNYKA 434
Db 260 FYYGKGLINVEAA 272
```

RESULT 11  
US-11-452-695-2  
/ Sequence 2, Application US/11452695  
/ Publication No. US20060222641A1

```
/ GENERAL INFORMATION:
/ APPLICANT: Neil H. Riordan
/ TITLE OF INVENTION: METHOD AND COMPOSITION FOR PREVENTING OR
/ TITLE OF INVENTION: REDUCING DEEP VEIN THROMBOSIS AND/OR PULMONARY EMBOLISM
/ FILE REFERENCE: AIDAN.005DV1
/ CURRENT APPLICATION NUMBER: US/11/452,695
/ CURRENT FILING DATE: 2006-06-14
/ PRIOR APPLICATION NUMBER: 60/468948
/ PRIOR FILING DATE: 2003-05-07
/ PRIOR APPLICATION NUMBER: 10/647131
/ PRIOR FILING DATE: 2003-08-22
/ NUMBER OF SEQ ID NOS: 7
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 2
/ LENGTH: 381
/ TYPE: PRT
/ ORGANISM: Bacillus Subtilis
US-11-452-695-2
```

```
Query Match 15.0%; Score 516.5; DB 7; Length 381;
Best Local Similarity 38.2%; Pred. No. 5.3e-29;
Matches 139; Conservative 53; Mismatches 123; Indels 49; Gaps 13;

QY 78 YLRMGAQYKYSKIIIPAVAKIKARDLLIAGMIDTGYFGNTRVSGIKPIODYKYQVD 137
Db 58 VISEKGGKXQKQPKRYNAAAATDEXAVKEL-----KKDPSVAIVEDH-IAHE 105
QY 138 DATSV-----SOGIADVTWNSLGYDGSVNSQIYGVAPGAKLVGVKVLGADGSGSVT 193
Db 106 YASQVYGISQIAKAPL-HSQGTGSNNVAVADSGIDSHPL--NVAGGASFVSETN 162
QY 194 PYDD-QAGHGTAVAGTGSVNSQIYGVAPGAKLVGVKVLGADGSGSVTIIAGYDWTY 252
Db 163 PYDDGSSHGTHVAGTIAALNN-SIGVLGVAFASLAVKVLDTSGSQYSWIINGIEMAI 221
QY 253 QNKDKXGIRVYNLSLSSQSDGTDLSQAVNNMADAIYVCVAAAGNSG--NNTYVSGP 310
Db 222 SN---NMDVTNMSIG---GPTGSTALKVTYDRAVSSGIVVAAAGNSGSGSTIVGYP 274
QY 311 AASKVITVGVADSNDIASFSSRGPTADGRLKPEVAPGVDDIAPRASGTSMGTINDY 370
Db 275 AKPSTIAYGAVVSSNQRASFSSVGSGLD-----VMAPEVSIQSTLPST----- 319
QY 371 YTRASGTSMATPHYSGVGALLIOAHPSWTDPKVTALLETADIAPKEIADIIYAGRVN 430
Db 320 YGANGTSMATPHVAGAAALILSKHPTWTAQVRDLSTATNIGD-----SSFYGGLIN 375
QY 431 YTKA 434
Db 376 VQAA 379
```

RESULT 12  
US-11-452-695-3  
/ Sequence 3, Application US/11452695  
/ Publication No. US20060222641A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Neil H. Riordan  
/ TITLE OF INVENTION: METHOD AND COMPOSITION FOR PREVENTING OR  
/ TITLE OF INVENTION: REDUCING DEEP VEIN THROMBOSIS AND/OR PULMONARY EMBOLISM  
/ FILE REFERENCE: AIDAN.005DV1  
/ CURRENT APPLICATION NUMBER: US/11/452,695  
/ CURRENT FILING DATE: 2006-06-14  
/ PRIOR APPLICATION NUMBER: 60/468948  
/ PRIOR FILING DATE: 2003-05-07  
/ PRIOR APPLICATION NUMBER: 10/647131  
/ PRIOR FILING DATE: 2003-08-22  
/ NUMBER OF SEQ ID NOS: 7  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 3  
/ LENGTH: 381  
/ TYPE: PRT  
/ ORGANISM: Bacillus Subtilis

US-11-452-695-3

Query Match 14.9%; Score 513.5; DB 7; Length 381;  
 Best Local Similarity 38.2%; Pred. No. 8.7e-29;  
 Matches 139; Conservative 52; Mismatches 124; Indels 49; Gaps 13;

QY 78 VRLMGAQVRYKYKIPAVAVKIKARDLLIAGMIDTGFNGTRVSGIKFIQEDYKVOVD 137  
 DB 58 VISEKGGVQKQKPKYVNAATAATLDEKAVKEL-----KKDPSVAVVEBDH-IAHE 105  
 QY 138 DATSV-----SQTGADTVNMSLGYDGGVVVAIVDTGIDANHPDLKKGKVGWYDAVNGRST 193  
 DB 106 YAGSVYVYGISQIKAPAL-HSQGYTGSNVKVAVIDSGIDISHPDL--NVAGGASFVSEEN 162  
 QY 194 PYDD-OGHGTHTVAGIYAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGSVSTIAGVDMV 252  
 DB 163 PYDGGSHGTHTVAGTIALALNN-SIGVLGVAFPSASLVAVKVLDSTGSGQYSWIINGIEMAI 221  
 QY 253 QNRKXGIRVINLSLSSQSSDGTDSLGAANNAMDAIGVVCVAAGNSGP--NTYTVGSP 310  
 DB 222 SN-----NMDVIMNSLG---GPTGSTALKTVVDKAVSSGIVVAAAAGNESSGSTSTVGY 274  
 QY 311 AAASKYITVGAVDSDNINIAFSRRGPTADGRKPEVVAAGVDIIAPRASGTSMTPTINDY 370  
 DB 275 AKYPTIAGVAVNSNORASFSSVSGSELD-----VMAGVSIQSTLPGGT----- 319  
 QY 371 YTKASGTSWATPHVSGVGLLIIQAPSWTPDKVKYKTALIEADIAPKEIADIAYGAGRYN 430  
 DB 320 YGAVNGTSWATPHVAGAAALLISKHPTWNAOVRLESTAYTLG---NSFYGKGLIN 375  
 QY 431 YVKA 434  
 DB 376 VQAA 379

RESULT 13

US-11-452-695-4  
 ; Sequence 4, Application US/11452695  
 ; Publication No. US20060222641A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Neil H. Riordan  
 ; TITLE OF INVENTION: METHOD AND COMPOSITION FOR PREVENTING OR  
 ; TITLE OF INVENTION: REDUCING DEEP VEIN THROMBOSIS AND/OR PULMONARY EMBOLISM  
 ; FILE REFERENCE: AIDAN.005DV1  
 ; CURRENT APPLICATION NUMBER: US/11/452,695  
 ; PRIOR FILING DATE: 2006-06-14  
 ; PRIOR APPLICATION NUMBER: 60/468948  
 ; PRIOR FILING DATE: 2003-05-07  
 ; PRIOR APPLICATION NUMBER: 10/647131  
 ; PRIOR FILING DATE: 2003-08-22  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 381  
 ; TYPE: PRT  
 ; ORGANISM: Bacillus Subtilis  
 US-11-452-695-4

Query Match 14.9%; Score 513.5; DB 7; Length 381;  
 Best Local Similarity 38.2%; Pred. No. 8.7e-29;  
 Matches 139; Conservative 52; Mismatches 124; Indels 49; Gaps 13;

QY 78 VRLMGAQVRYKYKIPAVAVKIKARDLLIAGMIDTGFNGTRVSGIKFIQEDYKVOVD 137  
 DB 58 VISEKGGVQKQKPKYVNAATAATLDEKAVKEL-----KKDPSVAVVEBDH-IAHE 105  
 QY 138 DATSV-----SQTGADTVNMSLGYDGGVVVAIVDTGIDANHPDLKKGKVGWYDAVNGRST 193  
 DB 106 YAGSVYVYGISQIKAPAL-HSQGYTGSNVKVAVIDSGIDISHPDL--NVAGGASFVSEEN 162  
 QY 194 PYDD-OGHGTHTVAGIYAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGSVSTIAGVDMV 252  
 DB 163 PYDGGSHGTHTVAGTIALALNN-SIGVLGVAFPSASLVAVKVLDSTGSGQYSWIINGIEMAI 221

QY 253 QNRKXGIRVINLSLSSQSSDGTDSLGAANNAMDAIGVVCVAAGNSGP--NTYTVGSP 310  
 DB 222 SN-----NMDVIMNSLG---GPTGSTALKTVVDKAVSSGIVVAAAAGNESSGSTSTVGY 274  
 QY 311 AAASKYITVGAVDSDNINIAFSRRGPTADGRKPEVVAAGVDIIAPRASGTSMTPTINDY 370  
 DB 275 AKYPTIAGVAVNSNORASFSSVSGSELD-----VMAGVSIQSTLPGGT----- 319  
 QY 371 YTKASGTSWATPHVSGVGLLIIQAPSWTPDKVKYKTALIEADIAPKEIADIAYGAGRYN 430  
 DB 320 YGAVNGTSWATPHVAGAAALLISKHPTWNAOVRLESTAYTLG---NSFYGKGLIN 375  
 QY 431 YVKA 434  
 DB 376 VQAA 379

RESULT 14

US-10-526-324-20  
 ; Sequence 20, Application US/10526324  
 ; Publication No. US20060248617A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Imanaka, Takayuki  
 ; APPLICANT: Atom, Haruyuki  
 ; TITLE OF INVENTION: METHOD OF TARGETED GENE DISRUPTION, GENOME OF  
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE BACTERIUM AND GENOME CHIP USING  
 ; FILE REFERENCE: THE SAME  
 ; FILE REFERENCE: 490051.401USPC  
 ; CURRENT APPLICATION NUMBER: US/10/526,324  
 ; CURRENT FILING DATE: 2005-02-28  
 ; PRIOR APPLICATION NUMBER: PCT/IB2003/003597  
 ; PRIOR FILING DATE: 2003-08-29  
 ; PRIOR APPLICATION NUMBER: JP 2002-319011  
 ; PRIOR FILING DATE: 2002-08-30  
 ; NUMBER OF SEQ ID NOS: 2167  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 20  
 ; LENGTH: 524  
 ; TYPE: PRT  
 ; ORGANISM: Thermococcus kodakaraensis KOD1  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (414542)..(414542)  
 ; OTHER INFORMATION: n is a or c or g or t.  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (786890)..(786890)  
 ; OTHER INFORMATION: n is a or c or g or t.  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (786907)..(786907)  
 ; OTHER INFORMATION: n is a or c or g or t.  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (786946)  
 ; OTHER INFORMATION: n is a or c or g or t.  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (839139)..(839139)  
 ; OTHER INFORMATION: n is a or c or g or t.  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1128488)  
 ; OTHER INFORMATION: n is a or c or g or t.  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1128499)  
 ; OTHER INFORMATION: n is a or c or g or t.  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1128505)  
 ; OTHER INFORMATION: n is a or c or g or t.







```

QY 1 MKGLKALLIIVILVGLVGVSAAPAAPEKKVBOVRNVEKNYGLTPGLFRKIQKLNPNBEIS 60
DB 1 MKGLKALLIIVILVGLVGVSAAPAAPEKKVBOVRNVEKNYGLTPGLFRKIQKLNPNBEIS 60
QY 61 TVIVFENHREKEIAVRVLELMGAKRVYVHHIIPALADLVKRDLLVLSGLTGKAKLSGV 120
DB 61 TVIVFENHREKEIAVRVLELMGAKRVYVHHIIPALADLVKRDLLVLSGLTGKAKLSGV 120
QY 121 RFIQEDKYKTVSALEGLDESAAQVMATYVMNLGYDSSGITTIGIDTGDASHPLDQKV 180
DB 121 RFIQEDKYKTVSALEGLDESAAQVMATYVMNLGYDSSGITTIGIDTGDASHPLDQKV 180
QY 181 IGMWDFVNGRSYFYDDHGHGTHVASIAAGTGAASNGKYKMGAPGAKLAGIKVLGADSGS 240
DB 181 IGMWDFVNGRSYFYDDHGHGTHVASIAAGTGAASNGKYKMGAPGAKLAGIKVLGADSGS 240
QY 241 ISTTIKGVMAVDNKKYGIKVINLSLSSQSSDGTDSLQAVNNAMDAGIIVCVAAAGNS 300
DB 241 ISTTIKGVMAVDNKKYGIKVINLSLSSQSSDGTDSLQAVNNAMDAGIIVCVAAAGNS 300
QY 301 GPMYTVGSPAAASKVITVGAVDNDNIASFSSRGPTADGRLEKPEVVAPEGVDIIAPRASG 360
DB 301 GPMYTVGSPAAASKVITVGAVDNDNIASFSSRGPTADGRLEKPEVVAPEGVDIIAPRASG 360
QY 361 TSMGTPINDYTTKASGTSMAATPHVSGVALLIOAHPSWTPDKYKTALISTADIAPKEIA 420
DB 361 TSMGTPINDYTTKASGTSMAATPHVSGVALLIOAHPSWTPDKYKTALISTADIAPKEIA 420
QY 421 DIAVAGRVNVYKAIKYDDYAKLTFGTSVADKGSATHTFVSGATFTVATLWYDTSDDI 480
DB 421 DIAVAGRVNVYKAIKYDDYAKLTFGTSVADKGSATHTFVSGATFTVATLWYDTSDDI 480
QY 481 DLVLYDPNGNEVDYSYATAYGFEKVGYNPTAGTWTVKVSYGKAANYQVDVSDSLQ 540
DB 481 DLVLYDPNGNEVDYSYATAYGFEKVGYNPTAGTWTVKVSYGKAANYQVDVSDSLQ 540
QY 541 SGGGNPNPNPNPTPTTDTOTFTGSVNDYMDSDTFMNMNSGATKITGDLTFDTSYND 600
DB 541 SGGGNPNPNPNPTPTTDTOTFTGSVNDYMDSDTFMNMNSGATKITGDLTFDTSYND 600
QY 601 LDLYLDPNGNLVDRSTSSNSYEHVEYANPAPGTMFLVYASTYGADYQLKAVVYGG 659
DB 601 LDLYLDPNGNLVDRSTSSNSYEHVEYANPAPGTMFLVYASTYGADYQLKAVVYGG 659

```

RESULT 2  
ID AAM24129 standard; protein; 654 AA.

```

XX AAM24129;
XX 17-OCT-2003 (revised)
XX 20-APR-1998 (first entry)
XX Pyrococcus furiosus protease PFUS.
XX Protease; research reagent; thermal stability; pyrococcus furiosus.
XX Pyrococcus furiosus; DSM-3638.
XX MO9721823-A1.
XX 19-JUN-1997.
XX 07-NOV-1996; 96MO-JP003253.
XX 12-DEC-1995; 95JP-00323285.
XX (TAKI ) TAKARA SHUZO CO LTD.
XX Takakura H, Morishita M, Yamamoto K, Mita M, Asada K;
XX Tsunashawa S, Kato I;
PI

```

```

XX WPI: 1997-332794/30.
DR N-PSDB; AAT85635.
XX Protease(s) and genes encoding them obtained from Thermococcus and
PT Pyrococcus strains - have extremely high thermal stability and are useful
PT industrially and as research reagents.
XX Disclosure; Page 125-128; 159pp; Japanese.
XX This sequence represents the protease from Pyrococcus furiosus DSM-3638.
CC This sequence is a protease of the invention. The proteases of the
CC invention have extremely high thermal stability. The proteases can be
CC used as research reagents, and industrially in the food, drug and
CC chemical industries. (updated on 17-OCT-2003 to standardise Os field)
XX
SQ Sequence 654 AA;
Query Match 87.4%; Score 2997.5; DB 2; Length 654;
Best Local Similarity 88.3%; Pred. No. 2.3e-181;
Matches 582; Conservative 22; Mismatches 50; Indels 5; Gaps 1;

```

```

QY 1 MKGLKALLIIVILVGLVGVSAAPAAPEKKVBOVRNVEKNYGLTPGLFRKIQKLNPNBEIS 60
DB 1 MKGLKALLIIVILVGLVGVSAAPAAPEKKVBOVRNVEKNYGLTPGLFRKIQKLNPNBEIS 60
QY 61 TVIVFENHREKEIAVRVLELMGAKRVYVHHIIPALADLVKRDLLVLSGLTGKAKLSGV 120
DB 61 TVIVFENHREKEIAVRVLELMGAKRVYVHHIIPALADLVKRDLLVLSGLTGKAKLSGV 120
QY 121 RFIQEDKYKTVSALEGLDESAAQVMATYVMNLGYDSSGITTIGIDTGDASHPLDQKV 180
DB 121 RFIQEDKYKTVSALEGLDESAAQVMATYVMNLGYDSSGITTIGIDTGDASHPLDQKV 180
QY 181 IGMWDFVNGRSYFYDDHGHGTHVASIAAGTGAASNGKYKMGAPGAKLAGIKVLGADSGS 240
DB 181 IGMWDFVNGRSYFYDDHGHGTHVASIAAGTGAASNGKYKMGAPGAKLAGIKVLGADSGS 240
QY 241 ISTTIKGVMAVDNKKYGIKVINLSLSSQSSDGTDSLQAVNNAMDAGIIVCVAAAGNS 300
DB 241 ISTTIKGVMAVDNKKYGIKVINLSLSSQSSDGTDSLQAVNNAMDAGIIVCVAAAGNS 300
QY 301 GPMYTVGSPAAASKVITVGAVDNDNIASFSSRGPTADGRLEKPEVVAPEGVDIIAPRASG 360
DB 301 GPMYTVGSPAAASKVITVGAVDNDNIASFSSRGPTADGRLEKPEVVAPEGVDIIAPRASG 360
QY 361 TSMGTPINDYTTKASGTSMAATPHVSGVALLIOAHPSWTPDKYKTALISTADIAPKEIA 420
DB 361 TSMGTPINDYTTKASGTSMAATPHVSGVALLIOAHPSWTPDKYKTALISTADIAPKEIA 420
QY 421 DIAVAGRVNVYKAIKYDDYAKLTFGTSVADKGSATHTFVSGATFTVATLWYDTSDDI 480
DB 421 DIAVAGRVNVYKAIKYDDYAKLTFGTSVADKGSATHTFVSGATFTVATLWYDTSDDI 480
QY 481 DLVLYDPNGNEVDYSYATAYGFEKVGYNPTAGTWTVKVSYGKAANYQVDVSDSLQ 540
DB 481 DLVLYDPNGNEVDYSYATAYGFEKVGYNPTAGTWTVKVSYGKAANYQVDVSDSLQ 540
QY 541 SGGGNPNPNPNPTPTTDTOTFTGSVNDYMDSDTFMNMNSGATKITGDLTFDTSYND 600
DB 541 SGGGNPNPNPNPTPTTDTOTFTGSVNDYMDSDTFMNMNSGATKITGDLTFDTSYND 600
QY 601 LDLYLDPNGNLVDRSTSSNSYEHVEYANPAPGTMFLVYASTYGADYQLKAVVYGG 659
DB 596 LDLYLDPNGNLVDRSTSSNSYEHVEYANPAPGTMFLVYASTYGADYQLKAVVYGG 654

```

RESULT 3  
ID AAM94841 standard; protein; 654 AA.

```

XX AAM94841;
XX

```

DT 06-MAY-1999 (first entry)  
 XX Hyperthermostable protease.  
 DE Hyperthermostable protease.  
 XX Hyperthermostable; protease; thermophilic; bacterium; subtilisin;  
 KM additive; drug; washing agent; foodstuff; chemical synthesis.  
 XX Pyrococcus furiosus.  
 OS MO9856926-A1.  
 XX  
 PN 17-DEC-1998.  
 PD  
 XX 04-JUN-1998; 98WO-JP002465.  
 PF  
 XX 10-JUN-1997; 97JP-00151969.  
 PR  
 XX (TAKI ) TAKARA SHUZO CO LTD.  
 PA  
 XX Takakura H, Morishita M, Shimojo T, Asada K, Kato I;  
 PI  
 XX WPI: 1999-080907/07.  
 DR N-PSDB; AAX05929.  
 DR  
 XX Recombinant hyperthermostable protease from *Pyrococcus furiosus* - and  
 PT gene encoding it, for large scale production of the protease for  
 PT industrial use.  
 XX  
 PS Disclosure; Page 60-63; 82pp; Japanese.  
 XX  
 XX The invention relates to a hyperthermostable protease derived from a  
 CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has  
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10  
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at  
 CC 95 deg.C. The invention also provides gene sequences encoding a  
 CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal  
 CC peptide from subtilisin, and PRO is the above protease. Host cells  
 CC (especially *Bacillus* strains) transformed with vectors comprising the  
 CC genes are used for the recombinant production of the protease. The  
 CC hyperthermostable protease which can be prepared in quantity suitable for  
 CC industrial use, can be used as an additive for drugs, washing agents and  
 CC foodstuffs and for chemical synthesis  
 CC  
 XX Sequence 654 AA:  
 SQ  
 Query Match 87.4%; Score 2997.5; DB 2; Length 654;  
 Best Local Similarity 88.3%; Pred. No. 2.3e-181;  
 Matches 582; Conservative 22; Mismatches 50; Indels 5; Gaps 1;

QY 361 TSMGTPINDYTKASGTSMTATPHVSGVALLIQAPHSWTPDKYKTALETADIIVAPKEIA 420  
 DB 361 TSMGTPINDYTKASGTSMTATPHVSGVALLIQAPHSWTPDKYKTALETADIIVAPKEIA 420  
 QY 421 DIAYGAGRNVYKAIKIDYALKTFTGSVADGSAHTFTDVGATVNTLTWDTGSSDI 480  
 DB 421 DIAYGAGRNVYKAIKIDYALKTFTGSVADGSAHTFTDVGATVNTLTWDTGSSDI 480  
 QY 481 DLYLDPNNGNEVDYATAYGFEKGYNPTAGTWTVKVSGAANYOVVYSDGSLQ 540  
 DB 481 DLYLDPNNGNEVDYATAYGFEKGYNPTAGTWTVKVSGAANYOVVYSDGSLQ 540  
 QY 541 SGGGNPNPNPNPTPTDTGTFTGSVNDYMDTSDTFNNVNSGATKIGDLTFDTSYND 600  
 DB 541 SGGGNPNPNPNPTPTDTGTFTGSVNDYMDTSDTFNNVNSGATKIGDLTFDTSYND 600  
 QY 601 LDLYLDPNNGNVDNSTSNSTYEHYANPACTMTFLYVASTYGMADYOLKAVYYG 659  
 DB 596 LDLYLDPNNGNVDNSTSNSTYEHYANPACTMTFLYVASTYGMADYOLKAVYYG 659  
 RESULT 4  
 AAW24121  
 ID AAW24121 standard; protein; 659 AA.  
 XX  
 AC AAW24121;  
 XX  
 DT 17-OCT-2003 (revised)  
 DT 20-APR-1998 (first entry)  
 XX  
 DE Thermococcus protease.  
 XX  
 KM Protease; research reagent; thermal stability; thermococcus celer.  
 OS Thermococcus celer; DSM-2476.  
 XX  
 PN WO9721823-A1.  
 XX  
 PD 19-JUN-1997.  
 XX  
 XX 07-NOV-1996; 96WO-JP003253.  
 PF  
 XX 12-DEC-1995; 95JP-00323285.  
 PR  
 XX (TAKI ) TAKARA SHUZO CO LTD.  
 PA  
 XX Takakura H, Morishita M, Yamamoto K, Mita M, Asada K;  
 PI Teunasaawa S, Kato I;  
 DR WPI: 1997-332794/30.  
 DR N-PSDB; AAT85667.  
 DR  
 XX Protease(s) and genes encoding them obtained from *Thermococcus* and  
 PT *Pyrococcus* strains - have extremely high thermal stability and are useful  
 PT industrially and as research reagents.  
 XX  
 PS Claim 1; Page 82-85; 159pp; Japanese.  
 XX  
 XX This sequence represents the protease from *Thermococcus celer* DSM-2476.  
 CC This sequence is a protease of the invention. The proteases of the  
 CC invention have extremely high thermal stability. The proteases can be  
 CC used as research reagents, and industrially in the food, drug and  
 CC chemical industries. (Updated on 17-OCT-2003 to standardise OS field)  
 CC  
 XX Sequence 659 AA;  
 SQ  
 Query Match 85.0%; Score 2914; DB 2; Length 659;  
 Best Local Similarity 83.7%; Pred. No. 4.5e-176;  
 Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps 7;

|          |  |  |     |
|----------|--|--|-----|
| QY       | 56   | ELSTVYFENHNEKEIAVAVLLEMGAKVRVYHIIIPALADLKVRDLIVTSGL--TG--G     | 113 |
| Db       | 59   | EDVTYMFESYGRDRAVKVLRIMGAQVKSYKIPAVAVKIKARDLLIAGMIDYFG          | 118 |
| QY       | 114  | KAKLSGVRPIQEDYKVTVALEGLDESAQAQVATYVMN-IGYDGSGITITGIDTGIDAS     | 172 |
| Db       | 119  | NTRVSGIKFIQEDYKQVQVDA-----TSVQIGADYVMNLSIGDGSQVVAIVDTGIDAN     | 173 |
| QY       | 173  | HPDLQKRTIGVWDVFNNGRSTPYDDHGHGHTVVASIAAGTGAASNGTKYKMAPGAKLAGIKV | 232 |
| Db       | 174  | HPDLKCKRTIGVWDVAVNGRSTPYDDQGHGHTVAGIVAGTGSV-NSQYIGVAPGAKLVGVKV | 232 |
| QY       | 233  | LGADSGSGISITIIKGYEMAVDNDKTXGIRVYNLSLSSSOSDQDTSLSQAIVNNAMAGIV   | 292 |
| Db       | 233  | LGADSGSGVSTIIAGVDVYVQNKDKTIGIRVYNLSLSSSOSDQDTSLSQAIVNNAMAGIV   | 292 |
| QY       | 293  | VCVAAGNSGPNITYVSGSPAASKVITVGAVDSNDNIASFSSRGPTADGRLKEPVAAGVD    | 352 |
| Db       | 293  | VCVAAGNSGPNITYVSGSPAASKVITVGAVDSNDNIASFSSRGPTADGRLKEPVAAGVD    | 352 |
| QY       | 353  | IIAPRASGTSMTGPIINDYTTKASGTSMTAPRHSVGCALILQAHPSWTPDKVTALLETAD   | 412 |
| Db       | 353  | IIAPRASGTSMTGPIINDYTTKASGTSMTAPRHSVGCALILQAHPSWTPDKVTALLETAD   | 412 |
| QY       | 413  | IVAPKEIIDIAAGAGRVNYYKAIKXDDYAKLPTFGSVADKSGATHEFDVSGATFTVATLY   | 472 |
| Db       | 413  | IVAPKEIIDIAAGAGRVNYYKAIKXDDYAKLPTFGSVADKSGATHEFDVSGATFTVATLY   | 472 |
| QY       | 473  | WDTGSDDIDLXYLDPRNGNEVDYSYTAAYGFEKGYGNPFPAGTWTVCVSGYKGAANYQVDV  | 532 |
| Db       | 473  | WDTGSDDIDLXYLDPRNGNEVDYSYTAAYGFEKGYGNPFPAGTWTVCVSGYKGAANYQVDV  | 532 |
| QY       | 533  | VSDGSLSSQGGGNPNPNPNPTPTTDTQTFGSGVNDYWDTSOTFTMANVNSGATKITGDL    | 592 |
| Db       | 533  | VSDGSLSSQGGGNPNPNPNPTPTTDTQTFGSGVNDYWDTSOTFTMANVNSGATKITGDL    | 592 |
| QY       | 593  | TFDTSYNDLIDLXYLPNGNLVDRSTSSNGBEHAYEYANPAPGTMFTLVVAYSTYGMADYOL  | 652 |
| Db       | 593  | TFDTSYNDLIDLXYLPNGNLVDRSTSSNGBEHAYEYANPAPGTMFTLVVAYSTYGMADYOL  | 652 |
| QY       | 653  | KAVVYYG 659  |     |
| Db       | 653  | KAVVYYG 659  |     |
| RESULT 5 |  |  |     |
| AAW94840 |  |  |     |
| ID       | AAW94840   | standard; protein; 659 AA.                                     |     |
| XX       | AAW94840;  |  |     |
| AC       | AAW94840;  |  |     |
| XX       | XX   |  |     |
| DT       | 06-MAY-1999  | (first entry)  |     |
| XX       | XX   |  |     |
| DE       | W098556926   | Seq ID 12.   |     |
| XX       | XX   |  |     |
| OS       | Thermococcus celer.                                  |  |     |
| XX       | XX   |  |     |
| PN       | W098556926-A1.                                       |  |     |
| PD       | 17-DEC-1998.   |  |     |
| PF       | 04-JUN-1998;   | 98WO-JP002465.   |     |
| XX       | XX   |  |     |
| PR       | 10-JUN-1997;   | 97JP-00151969.   |     |
| XX       | XX   |  |     |
| PA       | (TAKI ) TAKARA SHUZO CO LTD.                         |  |     |
| XX       | XX   |  |     |
| PI       | Takakura H, Morishita M, Shimojo T, Asada K, Kato I; |  |     |
| XX       | XX   |  |     |

|                           |  |                      |
|---------------------------|--|----------------------|
| DR                        |  | MP1: 1999-080907/07. |
| DR                        | N-PEDB; AA05926.   |                      |
| PT                        | Recombinant hyperthermostable protease from <i>Picrococcus furiosus</i> - and      |                      |
| PT                        | pI gene encoding it, for large scale production of the protease for                |                      |
| PT                        | industrial use.  |                      |
| PS                        | Disclosure; Page 55-58; 82pp; Japanese.  |                      |
| XX                        |  |                      |
| CC                        | The invention relates to a hyperthermostable protease derived from a               |                      |
| CC                        | thermophilic bacterium (especially <i>Picrococcus furiosus</i> ). The protease has |                      |
| CC                        | working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10            |                      |
| CC                        | (optimum 6-8), and retains more than 90% of its activity after 8 hours at          |                      |
| CC                        | 95 deg.C. The invention also provides gene sequences encoding a                    |                      |
| CC                        | polypeptide of formula SIG-Ala-Gly-Gly-Asn-Pro, where SIG is a signal              |                      |
| CC                        | peptide from subtilisin, and PRO is the above protease. Host cells                 |                      |
| CC                        | (especially <i>Bacillus</i> strains) transformed with vectors comprising the       |                      |
| CC                        | genes are used for the recombinant production of the protease. The                 |                      |
| CC                        | hyperthermostable protease which can be prepared in quantity suitable for          |                      |
| CC                        | industrial use, can be used as an additive for drugs, washing agents and           |                      |
| CC                        | foodstuffs and for chemical synthesis  |                      |
| XX                        |  |                      |
| SQ                        | Sequence 659 AA;   |                      |
|                           |  |                      |
| Query Match               | 85.0%; Score 2914; DB 2; Length 659;   |                      |
| Best Local Similarity     | 83.7%; Pred. No. 4,5e-176;   |                      |
| Matches 558; Conservative | 47; Mismatches 46; Indels 16; Gaps 77;   |                      |
| OY                        | 1 MKGLKALILVLLVGLGVGVAAPAEKKVBOVRN--VERNYGLITPGLFRKIQLKNPNE 57                     |                      |
| DB                        | 1 MKRIGAVALIALVLGLLALTAALAAPVKEV--VRNNAAVOCKNYGILLTPLLFKVKOHRMNO 58                |                      |
| OY                        | EISTVIYENHREKEIAVRVELMGAKRYVYHHIIIPATAOLKYRDLIVISGL--TG--G 113                     |                      |
| DB                        | 59 EVDIVIMGSIGDDRAVKULRMGAQVKYSYKIIIPAVAYIKARDLLIMGMDITGIFG 118                    |                      |
| OY                        | 114 KALSGVRFIOEDYKYTVASAELEGIDBSAQCVMATVVYN-LGYDGSGITIGIIDGTIDAS 172               |                      |
| DB                        | 119 NTRVSQIFIOEDYKQVDA-----TSVSDIGADVTWNSLGYGSGVVAIVDTGIDAN 173                    |                      |
| OY                        | 173 HPDILOKVIVGVNDVNRSPYRDDHGCHTHVASLAAGTGAASNQKRYGMAGATLAGIKY 232                 |                      |
| DB                        | 174 HPIDKGKITYGVDAVNGRSTPYDDQGHTHAAGIVAIGTSV-NSQYIGVAAGALKVGKV 232                 |                      |
| OY                        | 233 LGADGSGSITIIINGVEAVDNKDQYGIKVINLSLSSQSOSDGTDSLISQAANNAMDAGIV 292               |                      |
| DB                        | 233 LGADGSGSVSTIIINGVDMVGNKDKYGRIVINLSLSSQSOSDGTDSLISQAANNAMDAGIV 292              |                      |
| OY                        | 293 VCVAANGSGPNTTYVGSPPAAASKITIYGAVDSDNNIASFSSRGPTADGRLEPEVAPGDV 352               |                      |
| DB                        | 293 VCVAAANGSGPNTTYVGSPPAAASKITIYGAVDSDNNIASFSSRGPTADGRLEPEVAPGDV 352              |                      |
| OY                        | 353 IIAIRASGTSNGPTIINDYYTKASGTSMAATPVSGVGLLILOAHSMWPDPKKTALLITAD 412               |                      |
| DB                        | 353 IIAIRASGTSNGPTIINDYYTKASGTSMAATPVSGVGLLILOAHSMWPDPKKTALLITAD 412               |                      |
| OY                        | 413 IVAPKEIADIAYAGRAVNVYKAIKYDDYAKLTFTGSVADKGSATHTEPDVSGATEVTATLY 472              |                      |
| DB                        | 413 IVAPKEIADIAYAGRAVNVYKAIKYDDYAKLTFTGSVADKGSATHTEPDVSGATEVTATLY 472              |                      |
| OY                        | 473 WPDGSSSIDILYLDPNGNEVDYSTLAYVGERKGYVNPRTAGTYVYKVVSYKGAAANYQDV 532               |                      |
| DB                        | 473 WPDGSSSIDILYLDPNGNEVDYSTLAYVGEKGYVNPRTAGTYVYKVVSYKGAAANYQDV 532                |                      |
| OY                        | 533 VSDSLISQSGGGNPNPNPNPTTTDTQTGFSGVNDYMPQTSDFPTNNVNSGATKITGDL 592                 |                      |
| DB                        | 533 VSOGSLISQSGGGNPNPNPNPTTTDTQTGFSGVNDYMPQTSDFPTNNVNSGATKITGDL 592                |                      |
| OY                        | 593 TFDTSYNDLILYLDPNGNLVDRSTSNSEHEEYANPAEGTWELFYAYSTYGADYQL 652                    |                      |
| DB                        | 593 TFDTSYNDLILYLDPNGNLVDRSTSNSEHEEYANPAEGTWELFYAYSTYGADYQL 652                    |                      |
| OY                        | 653 KAVVYTG 659  |                      |

Db 653 KAVVYGG 659

RESULT 6  
ADN46373  
ID ADN46373 standard; protein; 663 AA.

AC ADN46373;

DT 01-JUL-2004 (first entry)

DE Thermococcus kodakaraensis KOD1 protein sequence SeqID251.

XX gene disruption; gene targeting; marker gene; transformation;  
XX homologous recombination; hyperthermostable archaeobacterium; KOD1;  
KW gene structure; gene function; enzyme activity; medicine;  
KW forensic science; food; drug inspection; molecular biology; immunology.  
XX Thermococcus kodakaraensis.

OS Thermococcus kodakaraensis.

PN WO2004022736-A1.

PD 18-MAR-2004.

PF 29-AUG-2003; 2003WO-IB003597.

PR 30-AUG-2002; 2002JP-00319011.

PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.

PI Imanaka T, Acomi H;

DR WPI; 2004-257583/24.

PT Method for disrupting targeted gene in genome of organism particularly  
PT thermostable bacterium and with genome chips for analysis, applicable in  
PT studying gene structure and functions.

PS Claim 9; SEQ ID NO 251; 598bp; Japanese.

XX This invention relates to a novel method for targeting disruption of an  
XX arbitrary gene in a genome of an organism which comprises providing the  
XX whole sequential data of the genome of such organism, selecting at least  
XX 1 arbitrary region in the sequence, providing a vector that contains a  
XX sequence homologous with the selected region and a marker gene,  
XX transformation, and homologous recombination. The genome is preferably  
XX the genome of a hyperthermostable archaeobacterium, particularly  
XX Thermococcus kodakaraensis KOD1. The method is for targeting the  
XX disruption of a gene in the genome of an organism, which is applicable in  
XX studying gene structure and functions as well as enzyme activities of  
XX encoded proteins and useful in medicine, forensic science, food or drug  
XX inspection, molecular biology and immunology. With this method, the  
XX disruption of a gene at an arbitrary position in a genome can be achieved  
XX efficiently and reliably. The present sequence is that of a protein  
XX encoded by the genome of Thermococcus kodakaraensis which was derived  
XX using the method of the invention. Note: The sequence data for this  
XX patent did not form part of the printed specification, but was obtained  
XX in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 663 AA;

Query Match 78.9%; Score 2705; DB 8; Length 663;

Best Local Similarity 77.7%; Pred. No. 7.9e-163;  
Matches 517; Conservative 60; Mismatches 80; Indels 8; Gaps 4;

QY 1 MKGLKLLIVIVLGIWGVSAAPAKKQVQVNV--EKRYGLITGGLPRKLOKLNPNRE 58

DB 1 MKKEGVALVLAFLVIGMAGSVLAPOKPA--VKNVQOQKRYGGLTGLFKVQVQKMSWDQ 58

QY 59 ISTVIVFENHREKEIVAVLELGAQVRYVYHIIIPAIADLKVRLVLSGL--TG--GK 114

Db 59 VSTIIFPNQADKEKAVEILDFLGAKKIKYNNHIIIPALAVKIKVKOLLIIAGLMDTGYFCN 118

QY 115 AKLSGVRPIQEDYKYTVSALEGLDESAAQVAVYVMNLGYDGSGLTIGIIDTGIDASHP 174

Db 119 AOLSGVQPIQEDYKYTVSALEGLDESAAQVAVYVMNLGYDGSGLTIGIIDTGIDASHP 178

QY 175 DLQKRVIGWVDFVNGRSYPPYDDHGHGTHVASIAAGTGAASNGKYGMAFGAKIAGIKVLG 234

Db 179 DLQKRVIGWVDFVNGRSYPPYDDHGHGTHVASIAAGTGAASNGKYGMAFGAKIAGIKVLN 238

QY 235 ADGSGSITIIKGVEMAYDNKRGYIKYINLSLSSQSSDGTDSLQAVNNAMADAGIYVC 294

Db 239 GQSSSISDIIINGVMAVQNKDKYGIKYNLSLSSQSSDGTDSLQAVNNAMADAGIYVV 298

QY 295 VAAGNSGPTTYVGSPPAASKVITGAVDSNINIASFSSRGPTAQRLEPVAAPGVDII 354

Db 299 VAAGNSGPKTYVGSPPAASKVITGAVDKYVITDFSSRGPTADNRLEPVAAPGVMII 358

QY 355 APRASGTSMTPIINDYTTKASGTSNATPHVSGVALLIQAPBSWTPDKVKTALIEADIV 414

Db 359 APRASGTSMTPIINDYTTKASGTSNATPHVAGIALLIQAPBSWTPDKVKTALIEADIV 418

QY 415 APKEIADIYAGARVNVYKAIKYDYAKLTFGSVADKGSATHTPDVSGATVATLYND 474

Db 419 KEDEIADIYAGARVNVYKAIKYDYAKLTFGSVADKGSATHTPDVSGATVATLYND 478

QY 475 TGSSDIDLTYLDPNGNEVDYSYATAYGFEKVGYYNPPTAGTMYVYSGYGAANYQVDVVS 534

Db 479 NSGSDIDLTYLDPNGNEVDYSYATAYGFEKVGYYNPPTAGTMYVYSGYGAANYQVDVVS 538

QY 535 DGLSLQSGGPNPNPNPNPTPTDTQTFGSDVNDYMDSDPTFMNNSGATKITGDLTF 594

Db 539 DGLSLQSGGSEPSPEPPTVDEKFTGTGVHVDYDSDPTFTMVNSGATKITGDLTF 598

QY 595 DTSYNDLILYDYPNGNLVDRSTSSNSYHVEYANPAPETWTFLYVAYSTYGADYQLKA 654

Db 599 DTSYNDLILYDYPNGNLVDRSESSNSYHVEYANPAPETWTFLYVAYDYGYADYQLDA 658

QY 655 VVYVG 659

Db 659 KVVYGG 663

RESULT 7

AAW24122  
ID AAW24122 standard; protein; 522 AA.

XX AAW24122;

DT 17-OCT-2003 (revised)

DT 20-APR-1998 (first entry)

XX Pyrococcus furiosus protease.

XX Protease; research reagent; thermal stability; pyrococcus furiosus.

XX Pyrococcus furiosus; DSM-3638.

XX Key Location/Qualifiers

FT Misc-difference 428 /label= Gly, Val

FT /note= "encoded by GNA"

PN WO9721823-A1.

PD 19-JUN-1997.

PF 07-NOV-1996; 96WO-JP003253.

PR 12-DEC-1995; 95JP-00323285.  
PA (TAKI ) TAKARA SHUZO CO LTD.

PI Takakura H, Morishita M, Yamamoto K, Mita M, Asada K;  
 PI Tsumasawa S, Kato I;  
 DR WPI, 1997-332794/30.  
 DR N-PSDB; AAT85668.  
 XX  
 PT Protease(s) and genes encoding them obtained from *Thermococcus* and  
 PT *Pyrococcus* strains - have extremely high thermal stability and are useful  
 XX industrially and as research reagents.  
 XX  
 PS Claim 5; Page 87-90; 159pp; Japanese.  
 XX  
 CC This sequence represents the protease from *Pyrococcus furiosus* DSM-3638.  
 CC This sequence is a protease of the invention. The proteases of the  
 CC invention have extremely high thermal stability. The proteases can be  
 CC used as research reagents, and industrially in the food, drug and  
 CC chemical industries. (Updated on 17-OCT-2003 to standardise OS field)  
 XX  
 SQ Sequence 522 AA;

Query Match 68.5%; Score 2349.5; DB 2; Length 522;  
 Best Local Similarity 85.2%; Pred. No. 2e-140;  
 Matches 449; Conservative 22; Mismatches 51; Indels 5; Gaps 1;

QY 133 AELGELDESSAQAQWATYVNNLGYDGSIGITIGITIGIDASHPDLQKVIQWDFNGRSY 192  
 DB 1 AELGELDESSAQAQWATYVNNLGYDGSIGITIGITIGIDASHPDLQKVIQWDFNGRSY 60  
 QY 193 PYDDHGHGHTVAAIAAGTGAASNGKYKMAPGAKIAGIKVLGADGSGSISTIIKGVMAV 252  
 DB 61 PYDDHGHGHTVAAIAAGTGAASNGKYKMAPGAKIAGIKVLGADGSGSISTIIKGVMAV 120  
 QY 253 DNKDKYGIKVINLSIGSSQSSDGTDSLQAVNNAMDAGIYVCAAGNSGPNYTVGSPAA 312  
 DB 121 DNKDKYGIKVINLSIGSSQSSDGTDSLQAVNNAMDAGIYVCAAGNSGPNYTVGSPAA 180  
 QY 313 ASKVIITVGAVDNNDNIASFSSRGPTADGRKPEVVAAGVDTIAPRASGTSMTGPIINDYTT 372  
 DB 181 ASKVIITVGAVDNNDNIASFSSRGPTADGRKPEVVAAGVDTIAPRASGTSMTGPIINDYTT 240  
 QY 373 KASGTSMATPHVSGVGLIIQAHPSWTPDKVKTALLETADIAPKEIADIAYGAGRNVY 432  
 DB 241 AAGTSMATPHVAGIAALLIQAHPSWTPDKVKTALLETADIAPKEIADIAYGAGRNVY 300  
 QY 433 KAIKYDYAKLTFTGSAADKGSATHTFDVSGATFTATLTYWDTGSSDIDLTYLDPNGNEV 492  
 DB 301 KAIYDNYAKLVFTGYVANKGSQTHQFVIGASFTVATLTYWDMANSDLDLYLDPNGNOV 360  
 QY 493 DYSTAYYGFKEKGYVNPATGTTVVKVSYKGAANYQVVDVSDGSLSSQSGGNGPNPNP 552  
 DB 361 DYSTAYYGFKEKGYVNPATGTTVVKVSYKGAANYQVVDVSDGSLSSQSGGNGPNPNP 415  
 QY 553 NPTFTDTQTFTGSVNDYWDTSDFTMNVNSGATKTIGDLTFDTSYNDLILYDPNGNU 612  
 DB 416 QPEPTVDAKTFQKSDHYVYRSDTFTMTVNSGATKTIGDLTFDTSYNDLILYDPNGNU 475  
 QY 613 VDRSTSSNVEHEVYANPARGTFTVLYAYSTYGADYQOLKAVVYTG 659  
 DB 476 VDRSSEPSNIEHVEYLTLPAGTMYFLVAYYTYGMAVEYELTAKVYTG 522

RESULT 8  
 AAM94838  
 ID AAM94838 standard; protein; 522 AA.  
 AC AAM94838;  
 XX  
 XX 06-MAY-1999 (first entry)  
 XX  
 XX Hyperthermostable protease.  
 XX  
 XX Hyperthermostable; protease; thermophilic; bacterium; subtilisin;  
 KW additive; drug; washing agent; foodstuff; chemical synthesis.

XX  
 OS *Pyrococcus furiosus*.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 428  
 FT /Label= Gly or Val  
 XX  
 PN MO9856926-A1.  
 XX  
 PD 17-DEC-1998.  
 XX  
 PF 04-JUN-1998; 98MO-JP002465.  
 XX  
 PR 10-JUN-1997; 97JP-00151969.  
 XX  
 PA (TAKI) TAKARA SHUZO CO LTD.  
 XX  
 PI Takakura H, Morishita M, Shimojo T, Asada K, Kato I;  
 DR WPI, 1999-080907/07.  
 XX  
 PT Recombinant hyperthermostable protease from *Pyrococcus furiosus* - and  
 PT gene encoding it, for large scale production of the protease for  
 PT industrial use.  
 XX  
 PS Claim 1; Page 39-41; 82pp; Japanese.

CC The invention relates to a hyperthermostable protease derived from a  
 CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has  
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10  
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at  
 CC 95 deg.C. The invention also provides gene sequences encoding a  
 CC polypeptide of formula SIG-ALA-Gly-Gly-Aasn-PRO, where SIG is a signal  
 CC peptide from subtilisin, and PRO is the above protease. Host cells  
 CC (especially *Bacillus* strains) transformed with vectors comprising the  
 CC genes are used for the recombinant production of the protease. The  
 CC hyperthermostable protease which can be prepared in quantity suitable for  
 CC industrial use, can be used as an additive for drugs, washing agents and  
 CC foodstuffs and for chemical synthesis

SQ Sequence 522 AA;

Query Match 68.5%; Score 2349.5; DB 2; Length 522;  
 Best Local Similarity 85.2%; Pred. No. 2e-140;  
 Matches 449; Conservative 22; Mismatches 51; Indels 5; Gaps 1;

QY 133 AELGELDESSAQAQWATYVNNLGYDGSIGITIGITIGIDASHPDLQKVIQWDFNGRSY 192  
 DB 1 AELGELDESSAQAQWATYVNNLGYDGSIGITIGITIGIDASHPDLQKVIQWDFNGRSY 60  
 QY 193 PYDDHGHGHTVAAIAAGTGAASNGKYKMAPGAKIAGIKVLGADGSGSISTIIKGVMAV 252  
 DB 61 PYDDHGHGHTVAAIAAGTGAASNGKYKMAPGAKIAGIKVLGADGSGSISTIIKGVMAV 120  
 QY 253 DNKDKYGIKVINLSIGSSQSSDGTDSLQAVNNAMDAGIYVCAAGNSGPNYTVGSPAA 312  
 DB 121 DNKDKYGIKVINLSIGSSQSSDGTDSLQAVNNAMDAGIYVCAAGNSGPNYTVGSPAA 180  
 QY 313 ASKVIITVGAVDNNDNIASFSSRGPTADGRKPEVVAAGVDTIAPRASGTSMTGPIINDYTT 372  
 DB 181 ASKVIITVGAVDNNDNIASFSSRGPTADGRKPEVVAAGVDTIAPRASGTSMTGPIINDYTT 240  
 QY 373 KASGTSMATPHVSGVGLIIQAHPSWTPDKVKTALLETADIAPKEIADIAYGAGRNVY 432  
 DB 241 AAGTSMATPHVAGIAALLIQAHPSWTPDKVKTALLETADIAPKEIADIAYGAGRNVY 300  
 QY 433 KAIKYDYAKLTFTGSAADKGSATHTFDVSGATFTATLTYWDTGSSDIDLTYLDPNGNEV 492  
 DB 301 KAIYDNYAKLVFTGYVANKGSQTHQFVIGASFTVATLTYWDMANSDLDLYLDPNGNOV 360  
 QY 493 DYSTAYYGFKEKGYVNPATGTTVVKVSYKGAANYQVVDVSDGSLSSQSGGNGPNPNP 552  
 DB 361 DYSTAYYGFKEKGYVNPATGTTVVKVSYKGAANYQVVDVSDGSLSSQSGGNGPNPNP 415

QY 553 NPTPTDFTGTSVNDYMDTSDTFMNVNSGATKITGDLTPTDYSNDLDLYLDPNGML 612  
 DB 416 QPEPTDARTFOKSDHYHYDRSDTFMTVNSGATKITGDLVPTSDHDDLXYIDPNOKL 475  
 QY 613 VDRSTSSNSYEHVEYANPAPGTWPLVYASTYGMADYOLKAVVYVG 659  
 DB 476 VDRSESPPNSYEHVEYLTLPAPGTWPLVYAYTYGMAVYELTAKVYVG 522

RESULT 9  
 AAM94836 standard; protein; 412 AA.  
 AAM94836;  
 06-MAY-1999 (first entry)  
 Hyperthermostable protease fragment.  
 Hyperthermostable; protease; thermophilic; bacterium; subtilisin;  
 additive; drug; washing agent; foodstuff; chemical synthesis.  
 Pyrococcus furiosus.  
 MO9856926-A1.  
 17-DEC-1998.  
 04-JUN-1998; 98WO-JP002465.  
 10-JUN-1997; 97JP-00151969.  
 (TAKI) TAKARA SHUZO CO LTD.  
 Takakura H, Morishita M, Shimojo T, Asada K, Kato I;  
 WPI: 1999-080907/07.  
 N-PSDB; AAX05920.  
 Recombinant hyperthermostable protease from *Pyrococcus furiosus* - and  
 gene encoding it, for large scale production of the protease for  
 industrial use.  
 Claim 2; Page 35-37; 82pp; Japanese.  
 The invention relates to a hyperthermostable protease derived from a  
 thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has  
 working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10  
 (optimum 6-8), and retains more than 90% of its activity after 8 hours at  
 95 deg.C. The invention also provides gene sequences encoding a  
 polypeptide of formula SIG-Ala-Gly-Gly-Asn-PRO, where SIG is a signal  
 peptide from subtilisin, and PRO is the above protease. Host cells  
 (especially *Bacillus* strains) transformed with vectors comprising the  
 genes are used for the recombinant production of the protease. The  
 hyperthermostable protease which can be prepared in quantity suitable for  
 industrial use, can be used as an additive for drugs, washing agents and  
 foodstuffs and for chemical synthesis

Sequence 412 AA;  
 Query Match 56.0%; Score 1918; DB 2; Length 412;  
 Best Local Similarity 89.5%; Pred. No. 3.3e-113;  
 Matches 367; Conservative 15; Mismatches 28; Indels 0; Gaps 0;

QY 133 AEELEGDESAAGVATYVWNLGYDGSGITIGITDGTIDASHPLDQKVGWDFVNGRSY 192  
 DB 1 AEELEGDESAAGVATYVWNLGYDGSGITIGITDGTIDASHPLDQKVGWDFVNGRSY 60  
 QY 193 PYDDHGHGTHVASIAAGTGAASNGKYKMAPGAKLAGIKVLGADGSGSISTIIKGVEMAV 252  
 DB 61 PYDDHGHGTHVASIAAGTGAASNGKYKMAPGAKLAGIKVLGADGSGSISTIIKGVEMAV 120

QY 253 DNKDKYGIKVINLSLSSQSSDSDGTDLSQAVNNAMADAGIVCAAGNSGPNTYVSSPA 312  
 DB 121 DNKDKYGIKVINLSLSSQSSDSDGTDLSQAVNNAMADAGIVCAAGNSGPNTYVSSPA 180  
 QY 313 ASKVITVGAVDSDNDINASTSSRGPTADGRLKEBVAPGVDDIIAPRASGTSMTGPIINDYTT 372  
 DB 181 ASKVITVGAVDKVDYTSFSSRGPTADGRLKEBVAPGVDDIIAPRASGTSMTGPIINDYTT 240  
 QY 373 KASGTSMATPHVSGVAGLILQAHPSWTPDKVKTALILETADIYAPKEIADIAGAGRVNY 432  
 DB 241 AAPGTSMATPHVAGLIALILQAHPSWTPDKVKTALILETADIYAPKEIADIAGAGRVNY 300  
 QY 433 KAIKYDDYAKLTFTGTSVADKGSATHTFDVSGATFVATLYMDTSSGSDLDLYLDPNGNEY 492  
 DB 301 KAINVDNNAKLVFTGVNANKSGQTHQFVIGASFTVATLYMDNANSDDLILYLDPNNGOY 360  
 QY 493 DYSYTAHYGFERYGYNPTAGTWTYKVVSYKGAANYQVDVSDGSLSSQSG 542  
 DB 361 DYSYTAHYGFERYGYNPTAGTWTYKVVSYKGAANYQVDVSDGSLSSQSG 410

RESULT 10  
 AAR87009 standard; protein; 237 AA.  
 AAR87009;  
 10-MAY-1996 (first entry)  
 Hyperthermostable protease.  
 Hyperthermostable; thermostability.  
 Pyrococcus furiosus.  
 Key Location/Qualifiers  
 FT Misc-difference 196  
 FT note="unspecified amino acid"  
 MO9534645-A1.  
 21-DEC-1995.  
 05-JUN-1995; 95WO-JP001095.  
 13-JUN-1994; 94JP-00130236.  
 26-JUL-1994; 94JP-00173912.  
 (TAKI) TAKARA SHUZO CO LTD.  
 Mitra M, Yamamoto K, Morishita M, Asada K, Tsunabawa S, Kato I;  
 WPI: 1996-049674/05.  
 N-PSDB; AAT08131.  
 The invention relates to the hyperthermostable protease of *Pyrococcus*  
 furiosus and its prodn. as a recombinant protein in transformants using a  
 vector carrying the protease gene (AAT08141). A genomic DNA sequence of  
 the invention is given in AAT08131 and its encoded protein in AAR87009

Sequence 237 AA;  
 Query Match 30.7%; Score 1051.5; DB 2; Length 237;  
 Best Local Similarity 87.9%; Pred. No. 1.3e-58;  
 Matches 203; Conservative 9; Mismatches 18; Indels 1; Gaps 1;

QY 175 DLQKVLGVWDFVNGRSYPYDDHGHGTHVASIAAGTGAASNGKYKMAPGAKLAGIKVLG 234  
 DB 175 DLQKVLGVWDFVNGRSYPYDDHGHGTHVASIAAGTGAASNGKYKMAPGAKLAGIKVLG 234

Db 1 DLKGVIGWVAVNKRSTPTDQGHGTHVAGIAGTGSV-NSQYIQAAPGAKLVGVKVLG 59  
 QY 235 ADGSGSITITIKGVEMAVDNKDKGIVKINLSLSSGSDGTDLSQAANNAMAGIIVC 294  
 Db 60 ADGSGSVSTIAGVDMVQNKDKGIVKINLSLSSGSDGTDLSQAANNAMAGIIVC 119  
 QY 295 VAAAGNSGPNNTVGSPPAASRVIVGAVDSNDNLSFSSSRGPTADGRKPEVVAPEGVDII 354  
 Db 120 VAAAGNSGPNNTVGSPPAASRVIVGAVDSNDNLSFSSSRGPTADGRKPEVVAPEGVDII 179  
 QY 355 APRAAGTSMGTPINDYYTKASGTSMATPVPVSGVALITIQAHPSWTPDKVKI 405  
 Db 180 APRAAGTSMGTPINDYYTKASGTSMATPVPVSGVALITIQAHPSWTPDKVKI 230

RESULT 11  
 ABB09483  
 ID ABB09483 standard; protein, 545 AA.  
 AC ABB09483;  
 XX  
 AC ABB09483;  
 XX  
 DT 11-SEP-2003 (revised)  
 DT 31-MAY-2002 (first entry)  
 XX  
 DE T. yonsei subtilisin-like serine protease.  
 XX  
 KM Subtilisin-like serine protease.  
 XX  
 OS Thermococcus yonseiensis.  
 XX  
 OS Thermococcus yonseiensis.  
 XX  
 PN KR2000072141-A.  
 XX  
 PD 05-DEC-2000.  
 XX  
 PF 04-AUG-2000; 2000KR-00045411.  
 XX  
 PR 04-AUG-2000; 2000KR-00045411.  
 XX  
 PA (KIMY/) KIM Y S.  
 XX  
 PI Chang HJ, Kim DH, Byun YR, Kim YS;  
 XX  
 DR WPI; 2001-298092/31.  
 DR N-PSDB; ABL54900.  
 XX  
 PT New DNA sequence of thermophilic protein decomposition enzyme and protein  
 PT derived therefrom.  
 XX  
 PS Claim 3; Page 8; 15pp; Korean.  
 PS  
 PS  
 CC This sequence represents the protein sequence of the Thermococcus  
 CC yonsei subtilisin-like serine protease of the invention. (Updated on 11-  
 CC SEP-2003 to standardise OS field)  
 CC  
 SQ Sequence 545 AA;

Query Match 26.5%; Score 909; DB 4; Length 545;  
 Best Local Similarity 44.4%; Pred. No. 4.2e-49;  
 Matches 213; Conservative 73; Mismatches 138; Indels 56; Gaps 14;

QY 36 EKNYGLTPTGKFKIKLPNERISTVIVFENREKEIARVLELGA-KVRVYHIIIPA 94  
 Db 3 DSKNKIPFDLEGRILNKPSEEFVITFNKPVSDADFTIAKNIGKINIKHRYKILPS 62  
 QY 95 IAAADLVKRVLLVSGITGKAKLSGVRFIOEDYKTVSALE---GLDESAAGVWATYV 150  
 Db 63 IAAADLVKRVLLVSGITGKAKLSGVRFIOEDYKTVSALE---GLDESAAGVWATYV 108  
 QY 151 KNLGYDSSGITTIGIDTGDASHPDQ-KGVIGWVDFVNGRSPYDDHGHGTHVASTAG 209  
 Db 109 -DFGVTKKITTALITDITGIDGNHVDLSGKTIQKDFINNKTPYDDNHGHGTHVASTAG 167  
 QY 210 TGAASNGKYGAPAKLAGIKVILGADGSGSISTITIKGVEMAVDNKDKGIVKINLSLGS 269

Db 168 TG-AGNSLYGVAPEALLVIGIKVLDANGSGSGMSTVAGIDMAVQNDQVGIKINISLGT 226  
 QY 270 SSSDGDPSLSQAANNAMAGIIVCVAAAGNSGNTTSSPPAASRVITVGAV---DSN 325  
 Db 227 STSSDGDPSLSQAANNAMAGIIVCVAAAGNSGNTTSSPPAASRVITVGAV---DSN 286  
 QY 326 DNIAFSSRGPTADGRKPEVVAPEGVDIIAPRASGTSMTGTPINDYYTKASGTSMATPVS 385  
 Db 287 DNIAFSSRGPTADGRKPEVVAPEGVDIIAPRASGTSMTGTPINDYYTKASGTSMATPVS 339  
 QY 386 GVCAILLQAHPSWTPDKVKITALEIADIVAPKEIADIVAGAGRVNYKAIKYDDVAKLTF 445  
 Db 340 GVCAILLQAHPSWTPDKVKITALEIADIVAPKEIADIVAGAGRVNYKAIKYDDVAKLTF 394  
 QY 446 TGSVAD-----KGSATTFPDVSGATF-VTATLV---WDTGSSPDILYLP 487  
 Db 395 KGNMIDVPHNYIYSGVLPGRSYSDTWTENATNTSYPIALITLIPDMANYNPDPDILYLP 454

RESULT 12  
 AAM24127  
 ID AAM24127 standard; protein, 168 AA.  
 AC AAM24127;  
 XX  
 AC AAM24127;  
 XX  
 DT 17-OCT-2003 (revised)  
 DT 20-APR-1998 (first entry)  
 XX  
 DE Pyrococcus protease fragment.  
 XX  
 KM Pyrococcus protease; research reagent; thermal stability; pyrococcus furiosus.  
 XX  
 OS Pyrococcus furiosus; DSM-3638.  
 XX  
 PN WO9721823-A1.  
 XX  
 PD 19-JUN-1997.  
 XX  
 PF 07-NOV-1996; 96WO-JP003253.  
 XX  
 PR 12-DEC-1995; 95JP-00323285.  
 XX  
 PA (TAKI) TAKARA SHUZO CO LTD.  
 XX  
 PI Takakura H, Morishita M, Yamamoto K, Mitra M, Asada K,  
 PI Terasawa S, Kato I;  
 XX  
 DR WPI; 1997-332794/30.  
 DR N-PSDB; AAV03761.  
 XX  
 PT Protease(s) and genes encoding them obtained from Thermococcus and  
 PT Pyrococcus strains - have extremely high thermal stability and are useful  
 PT industrially and as research reagents.  
 XX  
 PS Disclosure; Page 118-119; 159pp; Japanese.  
 PS  
 PS  
 CC This sequence represents a fragment of the protease from Pyrococcus  
 CC furiosus DSM-3638 (see AAM24127 for full length sequence). This sequence  
 CC is a fragment of the protease of the invention. The proteases of the  
 CC invention have extremely high thermal stability. The proteases can be  
 CC used as research reagents, and industrially in the food, drug and  
 CC chemical industries. (Updated on 17-OCT-2003 to standardise OS field)  
 CC  
 SQ Sequence 168 AA;

Query Match 25.2%; Score 864; DB 2; Length 168;  
 Best Local Similarity 100.0%; Pred. No. 6.6e-47;  
 Matches 168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 LVISGLTGKAKLSGVRFIOEDYKTVSALEGLDESAAGVWATYVNLGYDSSGITTIGI 164  
 Db 1 LVISGLTGKAKLSGVRFIOEDYKTVSALEGLDESAAGVWATYVNLGYDSSGITTIGI 60



QY 165 IDTGIDASHPDLOGKYIGWVDFVNGRSYPYDDHGHGTHVASIAGTGAASNGKYKGMARG 224  
 DB 61 IDTGIDASHPDLOGKYIGWVDFVNGRSYPYDDHGHGTHVASIAGTGAASNGKYKGMARG 120  
 QY 225 AKLAGIKVIGADSGSISTTIKGVEMAVNDKRYGIRVNLSTGSSQS 272  
 DB 121 AKLAGIKVIGADSGSISTTIKGVEMAVNDKRYGIRVNLSTGSSQS 168

## RESULT 13

AAR87011  
 ID AAR87011 standard; peptide; 188 AA.

XX AAR87011;

DT 10-MAY-1996 (first entry)

DE Peptide sequence.

XX Protease; hyperthermostable; thermostability.

OS Pyrococcus furiosus.

XX WO9534645-A1.

XX 21-DEC-1995.

PF 05-JUN-1995; 95WO-JP001095.

XX 13-JUN-1994; 94JP-00130236.

PR 26-JUL-1994; 94JP-00173912.

XX (TAKI ) TAKARA SHUZO CO LTD.

PI Mita M, Yamamoto K, Morishita M, Asada K, Tsunawawa S, Kato I;

DR WPI; 1996-049674/05.

XX Pyrococcus furiosus hyperthermostable protease gene - useful for

PT recombinant prodn. of hyperthermostable protease.

XX Example 4; Page 66-67; 85pp; Japanese.

CC The invention relates to a gene (AAT08141) that codes for a

CC hyperthermostable protease (AAR87007) of Pyrococcus furiosus. 2 DNA

CC sequences (AAT08133-34) are provided encoding peptides (AAR87010-11)

XX Sequence 188 AA;

Query Match 25.2%; Score 862.5; DB 2; Length 188;

Best Local Similarity 89.4%; Pred. No. 9.6e-47;

Matches 168; Conservative 8; Mismatches 11; Indels 1; Gaps 1;

QY 199 HGHVVASIAGTGAASNGKYKGMARGAKLAGIKVIGADSGSISTTIKGVEMAVNDKRY 258

DB 2 HGHVAGTYAGTGSV-NSQYIGVAPAKLVGVKVLGADSGSVSTIIAGVDWVYQNKDKY 60

QY 259 GIKVNLSTGSSQSDGTDLSQAANNAMDAIVCVAAGNSGPNITYVGSPPAAASKVIT 318

DB 61 GIKVNLSTGSSQSDGADLSQAANNAMDAIVCVAAGNSGPNITYVGSPPAAASKVIT 120

QY 319 VGAVDSNDNIASFSSGPTADGRLKPEVVAPEGVDIIAPRASGTSKGTPIINDYYTKASGTS 378

DB 121 VGAVDSNDNIASFSSGPTADGRLKPEVVAPEGVDIIAPRASGTSKGTPIINDYYTKASGTS 180

## RESULT 14

AAW24126

ID AAW24126 standard; protein; 188 AA.

XX AAW24126;

AC 17-OCT-2003 (revised)

DT 20-APR-1998 (first entry)

DE Thermococcus protease fragment.

XX Protease; research reagent; thermal stability; thermococcus celer.

OS Thermococcus celer; DSM-2476.

XX WO9721823-A1.

XX 19-JUN-1997.

PF 07-NOV-1996; 96WO-JP003253.

XX 12-DEC-1995; 95JP-00323285.

XX (TAKI ) TAKARA SHUZO CO LTD.

PI Takakura H, Morishita M, Yamamoto K, Mita M, Asada K;

DR WPI; 1997-332794/30.

DR N-PSDB; AAT85676.

XX Protease(s) and genes encoding them obtained from Thermococcus and

PT Pyrococcus strains - have extremely high thermal stability and are useful

XX industrially and as research reagents.

PS Disclosure; Page 110-112; 159pp; Japanese.

CC This sequence represents a fragment of the protease from Thermococcus

CC celer DSM-2476 (see AAW24121 for full length sequence). This sequence is

CC a fragment of the protease of the invention. The proteases of the

CC invention have extremely high thermal stability. The proteases can be

CC used as research reagents, and industrially in the food, drug and

CC chemical industries. (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 188 AA;

Query Match 25.2%; Score 862.5; DB 2; Length 188;

Best Local Similarity 89.4%; Pred. No. 9.6e-47;

Matches 168; Conservative 8; Mismatches 11; Indels 1; Gaps 1;

QY 199 HGHVVASIAGTGAASNGKYKGMARGAKLAGIKVIGADSGSISTTIKGVEMAVNDKRY 258

DB 2 HGHVAGTYAGTGSV-NSQYIGVAPAKLVGVKVLGADSGSVSTIIAGVDWVYQNKDKY 60

QY 259 GIKVNLSTGSSQSDGTDLSQAANNAMDAIVCVAAGNSGPNITYVGSPPAAASKVIT 318

DB 61 GIKVNLSTGSSQSDGADLSQAANNAMDAIVCVAAGNSGPNITYVGSPPAAASKVIT 120

QY 319 VGAVDSNDNIASFSSGPTADGRLKPEVVAPEGVDIIAPRASGTSKGTPIINDYYTKASGTS 378

DB 121 VGAVDSNDNIASFSSGPTADGRLKPEVVAPEGVDIIAPRASGTSKGTPIINDYYTKASGTS 180

## RESULT 15

AAW13667

AAW13667

ID AAW13667 standard; protein; 734 AA.

XX AAW13667;

DT 06-OCT-1997 (first entry)

DE Streptomyces viridoporus dhpa gene product.

XX asymmetric hydrolyase; dhpa; 4-substituted-1,4-dihydropyridine;  
 XX derivative; Streptomyces viridoporus; ester; chiral; synthesis;  
 KM cardiovascular; treatment; hypertension; ischaemic heart disease.  
 XX Streptomyces viridoporus.

XX Key Location/Qualifiers  
 FH Peptide 205..724  
 FT /note="fragment of dhpa protein; see AAM13666"

XX MO9705243-A1.

XX 13-FEB-1997.

XX 30-JUL-1996; 96WO-JP002147.

XX 31-JUL-1995; 95JP-00212975.

XX 29-FEB-1996; 96JP-00067478.

XX (SAOC) MERCIAN CORP.

XX Aritaawa A, Matsumuji M, Tsuruta T, Doehaei K, Nakashima T;  
 PI Isehiki K, Yoshiooka T;

XX WPI; 1997-145682/13.

XX N-PSDB; AAT61454.

XX Asymmetric hydrolyase gene derived from Streptomyces viridoporus - acts  
 PT on 4-substituted-1,4-dihydropyridine derivatives to produce chiral  
 PT derivatives useful for synthesis of cardiovascular drugs.

XX Claim 3; Page 49-55; 78pp; Japanese.

XX This sequence is an asymmetric hydrolase encoded by the Streptomyces  
 CC viridoporus dhpa gene. The enzyme acts on 4-substituted-1,4-  
 CC dihydropyridine derivatives. The enzyme allows the efficient conversion  
 CC of 4-substituted-1,4-dihydropyridine esters to chiral partially  
 CC hydrolysed derivatives, for use in the synthesis of cardiovascular drugs  
 CC suitable for the treatment of e.g. hypertension and ischaemic heart  
 CC disease

XX Sequence 734 AA;

Query Match 20.2%; Score 692.5; DB 2; Length 734;

Best Local Similarity 33.6%; Pred. No. 3.3e-35;

Matches 202; Conservative 62; Mismatches 223; Indels 115; Gaps 21;

QY 110 LTGGKAKLSVRIQEDYKVTVAELGLDESAQVMATYVWNLGYDGSGITTGIIIDTGI 169  
 DB 182 VTNGDRTASGIAHWMD-----GVRRAALDTSVGOIGAPKAMGAGYDGKVKIADVDTGV 236  
 QY 170 DASHPDLOGKIVGVFVNGRSYRYDDHGHGTHVAGTGAASNGKKGMAKAGLAG 229  
 DB 237 DTSHPDLOGKIVGVFVNGRSYRYDDHGHGTHVAGTGAASNGKKGMAKAGLAG 295  
 QY 230 IKVLGADSGSISTIIKGVMAVNDKKGIVKINLSGSOSSDGTDSLQAVNN-AMD 288  
 DB 296 GKVLDDSGFDDGGLAGMWA-----AQADVNNMSLGMDTPE-TDPLEAAVDKLSAE 350  
 QY 289 AGIVCVAAAGNGSPNTYVGSPPAAKVTIVGAVDSNDNIASFSSRGP-TADGRLKEPV 347  
 DB 351 KGVLFALAGNEPE--SIGSPGSAADALIVGAVDDKDLADFSSTGPRLDGALIKPDVT 408  
 QY 348 APCVDIAPRASGTSMTPIIND---YTTKASGTSMTPIVSGVGAALILQAPSWTPDKK 404  
 DB 409 ARGVDTTASABENDIGGVGEGPAGYMTISGTSMTPIVAGAAALLKQHPMTSABL 468  
 QY 405 TALISTADIAPREIADIYAGRVNYYAIK----- 436  
 DB 469 GAL--TGSTKGGK-YTPFEGSGRIQADKALQOTVLIADPVSVFQVQWPHHTDDEPVT 525

QY 437 -----YDDVAKLTFTGSVADKGSATHTFDVSGATFTATLYMDTGSSDIDLILYDPV 488  
 DB 526 LTRNIGTQDVDTKLTSTATDPRKGAAPAGFTLTGATTVVPA---GGSASVMTADTRL 582  
 QY 489 GNEVDYSYTAAYYGFKEKVGYNPPAGTWTVKVSY-----KGAANYQDVVSDGS 537  
 DB 583 GGTVDGAYSAYVAVATGGGQVTRTAAAVQREVEBVDVTRHIGRDKPTTEHLTDLIGVAG 642  
 QY 538 LSGSGGNRPNNPNTPTTDTQTF---TGS-VNDYMDTSDPTFM-----NNSGA 585  
 DB 643 LSGSGRYG-----APATDTATLRLPKGYTLVDSWIAKQFGLKGGIDMLVQPKLSV 693  
 QY 586 TKITGDLTPT-SYNDLIDLYLDPNGNLVDRSTSSSYEHVEYANPAGTWTFLVAYST 644  
 DB 694 TKDT-TLTIDARTTKAADITVPDPK-----AKPLSAT---IGTYDT 731  
 QY 645 YG 646  
 DB 732 AG 733

Search completed: January 6, 2007, 22:24:51  
 Job time: 146 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2007, 22:25:12 ; Search time 27 Seconds

(without alignments)  
2348.402 Million cell updates/sec

Title: US-10-800-684-5

Perfect score: 3428

Sequence: 1 MKGKALIVILVIGLAVGS.....YASTYGMADYQLKAVVYTG 659

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

PIR 80:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                |
|------------|-------|-------------|--------|-------|----------------------------|
| 1          | 688   | 20.1        | 442    | 2     | A69587 intracellular alka  |
| 2          | 667.5 | 19.5        | 444    | 2     | B83891 intracellular alka  |
| 3          | 555   | 16.2        | 382    | 1     | SUBSN subillisin (EC 3.4   |
| 4          | 554   | 16.2        | 379    | 1     | SUBSCL subillisin (EC 3.4  |
| 5          | 544   | 15.9        | 382    | 2     | I39780 subillisin (EC 3.4  |
| 6          | 540.5 | 15.8        | 1398   | 2     | T28159 pyrrolisin (EC 3.4  |
| 7          | 537.5 | 15.7        | 380    | 2     | A49778 high-alkaline seri  |
| 8          | 535   | 15.6        | 274    | 1     | SUBSD subillisin (EC 3.4   |
| 9          | 524   | 15.3        | 381    | 2     | JH0778 subillisin (EC 3.4  |
| 10         | 523   | 15.3        | 275    | 2     | JC1085 subillisin (EC 3.4  |
| 11         | 523   | 15.3        | 381    | 1     | SUBSI subillisin (EC 3.4   |
| 12         | 520   | 15.2        | 381    | 1     | SUBSS subillisin (EC 3.4   |
| 13         | 520   | 15.2        | 381    | 1     | JQ1487 subillisin (EC 3.4  |
| 14         | 508   | 14.8        | 361    | 2     | A48373 high-alkaline seri  |
| 15         | 508   | 14.8        | 361    | 2     | G83756 subillisin-type al  |
| 16         | 505.5 | 14.7        | 806    | 2     | A41341 microbial serine p  |
| 17         | 498.5 | 14.5        | 272    | 2     | A23624 subillisin (EC 3.4  |
| 18         | 496.5 | 14.5        | 384    | 2     | JC4802 alkaline proteinase |
| 19         | 496.5 | 14.5        | 757    | 2     | C84120 subillisin-type pr  |
| 20         | 496   | 14.5        | 374    | 2     | I39781 subillisin (EC 3.4  |
| 21         | 491.5 | 14.3        | 401    | 2     | I39974 serine proteinase   |
| 22         | 489.5 | 14.3        | 279    | 1     | SUMTV thermilase (EC 3.4   |
| 23         | 488   | 14.2        | 799    | 2     | G83753 subillisin-type pr  |
| 24         | 486.5 | 14.2        | 378    | 2     | A33973 high-alkaline seri  |
| 25         | 460   | 13.4        | 715    | 2     | JC4908 alkaline serine pr  |
| 26         | 452.5 | 13.2        | 1433   | 1     | A36734 bacillipeptidase F  |
| 27         | 451.5 | 13.2        | 372    | 2     | D83735 subillisin-type al  |
| 28         | 449   | 13.1        | 627    | 2     | D75393 serine proteinase,  |
| 29         | 446.5 | 13.0        | 535    | 2     | B82358 alkaline serine pr  |

|    |       |      |     |   |                            |
|----|-------|------|-----|---|----------------------------|
| 30 | 443   | 12.9 | 440 | 2 | H72784 probable alkaline   |
| 31 | 440   | 12.8 | 488 | 2 | A11930 proteinase (import  |
| 32 | 439.5 | 12.8 | 519 | 2 | S71451 haloalysin R4 (EC 3 |
| 33 | 437.5 | 12.8 | 397 | 2 | JM0075 cyteine-dependent   |
| 34 | 428.5 | 12.5 | 513 | 1 | A35742 agnalyisin (EC 3.4  |
| 35 | 426.5 | 12.4 | 534 | 1 | U50173 alkaline proteinase |
| 36 | 423   | 12.3 | 588 | 2 | C83836 subillisin-type pr  |
| 37 | 418   | 12.2 | 745 | 2 | JC6119 subillisin-type pr  |
| 38 | 417.5 | 12.2 | 401 | 2 | A57690 aerolysin precuro   |
| 39 | 417   | 12.2 | 747 | 2 | T06580 subillisin-like pr  |
| 40 | 416   | 12.1 | 645 | 1 | SUBSMP serine proteinase   |
| 41 | 409.5 | 11.9 | 525 | 2 | G84406 haloalysin (import  |
| 42 | 401.5 | 11.7 | 754 | 2 | T06579 subillisin-like pr  |
| 43 | 401   | 11.7 | 530 | 2 | A42605 haloalysin (EC 3.4  |
| 44 | 400   | 11.7 | 319 | 2 | I39866 microbial serine p  |
| 45 | 398   | 11.6 | 754 | 2 | T06577 subillisin-like pr  |

#### ALIGNMENTS

##### Result 1

A69587 intracellular alkaline serine proteinase aprx - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004

C:Accession: A69587

R:Kunet, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertec

A.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Gall

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.

Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauei,

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schoeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,

akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.

A:Authors: Yoshikawa, H.F.; Zumbstein, B.; Yoshikawa, H.; Zumbstein, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:980404033; PMID:9384377

A:Accession: A69587

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-442 <KUN>

A:Cross-references: UNIPROT:O31788; UNIPARC:UPI000006043F; GB:Z99113; GB:AL009126; NID:9;

A:Experimental source: strain 168

C:Genes:

A:Gene: aprx

F:146-398/Domain: subillisin homology <SBT>

Query Match

Best local similarity 39.7%; Pred. No. 1.8e-30; Mismatches 132; Indels 44; Gaps 14;

Matches 167; Conservative 78; Mismatches 132; Indels 44; Gaps 14;

|    |     |  |
|----|-----|--|
| QY | 43  | TFGLFRKI--OKLNDNEISTVIVFNHREK--EIAVRVLEL-MGAKRYVYHIIIPA 96     |
| DB | 35  | TFCLFHKFPETQLQNRKKMSVILIEFEGCHETGFQMGAGVILQEKRSKLKSFNKNCS 94   |
| QY | 97  | ADLKVRDLVLSGLTGKRAKSGVRFIOEDYVYVTSALIEGDESAQVM--ATVYVNLG- 154  |
| DB | 95  | AE-----VTPSALHSLSECSNIR-----FVYANREKVALDTRTESHAKEVVRNQ 142     |
| QY | 155 | -YDGSGITIGLIDTGDISHPDLQGVIGWDFVNGRSYFYDDHGHGTHVASTAAGTGA 213   |
| DB | 143 | TLTGKGVYAAVVDIGI-YPHPLBEGRIIGFADMVNQKIEPYDDNCHGTCAQDVASSGS 201 |
| QY | 214 | SNKTKGNAPGAKLGIKYLGDGSGSISTTIKGVEMAV---DNKDKGIVINLSLGS 269     |
| DB | 202 | SSGQYRGPAPFANLIGVVLNKGSGTGLADIEGVEMCIQYVNEPDPE-PIDIMNSLGS 260  |

```

Qy 270 S---QSSDGTSLQAVNNANMADAGIIVCAAGNSGPNITVYVSPAAASKVITTVGAVDSN- 325
Db 261 DALRYDHQEDPLVYAEASAGIIVCAAGNSGPDQITIASGVSKVITTVGALDDNN 320
Qy 326 -----DNIAFSSRGPTADRLKPEVVAPEVDIIAPASGT-----SMGTPINDYTTKA 374
Db 321 TASDDDTVASPSSRGPTVYGEKPEVDIIAPGVNIIISLSPSYIDKLQKSSRVSGVFTM 380
Qy 375 SGTSMATPHVSGVALLQAHPSWTPDKVTALLETADIYAPKEIADIYAGAGRVNYKA 434
Db 381 SGTSMATPHVSGVALLQAHPSWTPDKVTALLETADIYAPKEIADIYAGAGRVNYKA 438
Qy 435 I 435
Db 439 V 439

```

## RESULT 2

```

B83891
Intracellular alkaline serine proteinase aprX [imported] - Bacillus halodurans (strain C
C/Species: Bacillus halodurans
C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C/Accession: B83891
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A/Reference number: A83650; MUID:20512582; PMID:11058132
A/Accession: B83891
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-444 <SNO>
A/Cross-references: UNIPROT:Q9KBU7; UNIPARC:UPI000000C3D43; GB:AP001513; GB:BA000004; NIT
A/Experimental source: strain C-125
C/Genetics:
A/Gene: aprX

```

```

Query Match 19.5%; Score 667.5; DB 2; Length 444;
Best Local Similarity 36.7%; Pred. No. 2.3e-29;
Matches 156; Conservative 80; Mismatches 122; Indels 67; Gaps 12;

```

```

Qy 35 VERNYGLTTPGLFRKIQKLNPNNEISTVYFENHREKEIAVRVLELMGAKVRVYHIIIPA 94
Db 57 VEINOGCFDQ-----SMVYENHLKKGKQ-----COLRHSFKIRC 92
Qy 95 IAADL---KVRDLIVISGLTGAKAKLSGVRFIQEDYKVTVSALRG-LDESAQVWATYV 150
Db 93 VSNATVPESIQEMLVCKDIR-----KIYLRVHALDITAVESAQAPEV 136
Qy 151 WNLG--YDSGRTITGIIDPIDASHDLOGKVIGWDFVNGRSGYPYDDHGHGTHVASTA 208
Db 137 IRNGEITLTGKDVITIAVIDGI-YPHEDLEGRIKAFVDFVNGRSEPYDDHGHGTHCAGDA 195
Qy 209 GTGAASNGKYGAPAKLAGIKVLGADSGSISTIKGVEMAVDKDKY---GIKVINL 265
Db 196 GNGASSDGQYRGAPAPFANVYGVKLVKQNGSLSEIMQGVEMQIQNEHPDDPIHIIIM 255
Qy 266 SIGSS---QSSDGTSLQAVNNANMADAGIIVCAAGNSGPNITVYVSPAAASKVITTVGAV 322
Db 256 SIGGQALPYENEGEDPMVRIVEEAMNAGITVCAAGNSGPDQITIASPVSKEVITVGL 315
Qy 323 D-----GNDNIASFSSRGPTADRLKPEVVAPEVDIIAPASGT-----SMGTPINDY 370
Db 316 DDRDITDRDDDDVAPSSRGPTITYGKPEVDIIAPGVNIIISLSPSYIDKLQKSSRVSG 375
Qy 371 YTKASGTSMATPHVSGVALLQAHPSWTPDKVTALLETADIYAPKEIADIYAGAGRVN 430
Db 376 YTMMSGTSMATPHVSGVALLQAHPSWTPDKVTALLETADIYAPKEIADIYAGAGRVN 433
Qy 431 VKKAI 435
Db 434 AEGAI 438

```

## RESULT 3

SUBSN

subtilisin (EC 3.4.21.62) BPN' precursor - Bacillus amyloliquefaciens

N/Alternate names: subtilisin Novo

C/Species: Bacillus amyloliquefaciens

C/Date: 24-Apr-1984 #sequence\_revision 28-Aug-1985 #text\_change 05-Oct-2004

C/Accession: B25415; A93495; T44584; A92033; A00970

R/Vasantha, N.; Thompson, L.D.; Rhodes, C.; Banner, C.; Nagle, J.; Filipula, D.

J. Bacteriol. 159, 811-819, 1984

A/Title: Gene for alkaline protease and neutral proteases from Bacillus amyloliquefaciens

A/Reference number: A25415; MUID:85006739; PMID:6090391

A/Accession: B25415

A/Molecule type: DNA

A/Residues: 1-382 &lt;VAS&gt;

A/Cross-references: UNIPROT:P00782; UNIPARC:UPI0000136180; GB:K02496; NID:g142525; PIDN:f

A/Experimental source: ATCC 23844

R/Wells, J.A.; Ferrari, E.; Henner, D.J.; Batelli, D.A.; Chen, E.Y.

Nucleic Acids Res. 11, 7911-7925, 1983

A/Title: Cloning, sequencing, and secretion of Bacillus amyloliquefaciens subtilisin in f

A/Reference number: A93495; MUID:84069812; PMID:6316278

A/Accession: A93495

A/Molecule type: DNA

A/Residues: 1-382 &lt;WEL&gt;

A/Cross-references: UNIPARC:UPI0000136180

A/Accession: T44584

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 'W', 8-382 &lt;WEL&gt;

A/Cross-references: UNIPARC:UPI0000136180; EMBL:X00165; NID:g39337; PIDN:CAA2490.1; PID.

R/Marland, F.S.; Smith, E.L.

J. Biol. Chem. 242, 5198-5211, 1967

A/Title: Subtilisin BPN'. VII. Isolation of cyanogen bromide peptides and the complete an

A/Reference number: A92033; MUID:68086682; PMID:6065094

A/Accession: A92033

A/Molecule type: protein

A/Residues: 108-162, 'PN', 165-167, 'D', 169-194, 'SA', 197-204, 'DA', 207-264, 'ST', 267-357, 'Q', .

A/Cross-references: UNIPARC:UPI0000112C9B

R/Kraut, J.

in The Enzymes, 3rd ed., vol.3, Boyer, P.D., ed., pp.547-560, Academic Press, New York, 1

A/Title: Subtilisin: X-ray structure.

A/Reference number: A94443

A/Comments: annotation; X-ray crystallography, 2.5 angstroms; active site

C/Comment: Secretion of subtilisin is associated with the onset of sporulation, and many

not necessary for normal sporulation.

C/Genetics:

A/Start codon: GTG

C/Superfamily: Subtilisin, subtilisin homology

C/Keywords: hydrolase; serine proteinase

F/1-32/Domain: signal sequence #status predicted &lt;SIG&gt;

F/133-107/Domain: activation peptide #status predicted &lt;APT&gt;

F/108-382/Product: subtilisin BPN' #status experimental &lt;MPT&gt;

F/130-342/Domain: subtilisin homology &lt;SBN&gt;

F/139,171,328/Active site: Asp, His, Ser #status experimental

```

Query Match 16.2%; Score 555; DB 1; Length 382;
Best Local Similarity 35.1%; Pred. No. 2.7e-23;
Matches 158; Conservative 65; Mismatches 141; Indels 86; Gaps 16;

```

```

Qy 1 MKGKALIVLVIGLV---VGSVAAPPEKVVQVVRVENVNGYGLTTPGLFRKIQKLNPN 56
Db 1 MKGKALIVLVIGLV---VGSVAAPPEKVVQVVRVENVNGYGLTTPGLFRKIQKLNPN 56
Qy 57 EISTVYFENHREKEIAVR---VLELMGAKVRVYHIIIPAIDLK---VRDLIVISGL 110
Db 41 -----IVGFKQTMSTMSAAKKDVISERKGVQGFQKVVDAASATLNEKAVKL----- 89
Qy 111 TGGAKLSGVRFIEDYKVTVSALRGIDESAQVMAVYVWNLGYSGGITTIIGIDGID 170
Db 90 ----KQDPSVAVYVEDH--VAHAAYAGVPGVSGQIKAPALHSQGYTGSNVVAVVIDGID 143
Qy 171 ASHPDLOGKVIGWDFVNGRSGYPYDDHGHGTHVASTAAGTGAASNGK--YKGAAPAKL 227
Db 144 SSHPDL---KVAAGASNPVPSSTNPPQDNNSHGTHV---AGTVAALNNSIGVLGVAAPSAL 197

```

[illegible]

RESULT 4  
SUBJECT  
subtilisin (EC 3.4.21.62) Carlsberg precursor - *Bacillus licheniformis*  
C/Species: *Bacillus licheniformis*  
C/Date: 30-Jun-1998 #sequence\_revision 30-Jun-1998 #text\_change 05-Oct-2004  
C/Accession: A24111; A00968  
C/Accession: A24111; A00968  
J/Jacobs, M.; Eliasson, M.; Uhlen, M.; Flock, J.I.  
Nucleic Acids Res. 13, 6913-6926, 1985  
A/Title: Cloning, sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*  
A/Reference number: A24111; MUID:86093688; PMID:3001653  
A/Accession: A24111  
A/Molecule type: DNA  
A/Molecule type: DNA  
A/Residues: 1-379 <JUNC>  
A/Cross-references: UNIPROT:P00780, UNIPARC:UPI00000136181; GB:X03341; NID:g487721; PIDN  
A/Experimental source: Strain NCIB5816  
J/Smith, E.L.; Delange, R.U.; Evans, W.H.; Landon, M.; Markland, F.S.  
J. Biol. Chem. 243, 2184-2191, 1968  
A/Title: Subtilisin Carlsberg. V. The complete sequence; comparison with subtilisin BPN'  
A/Reference number: A00968; MUID:68234702; PMID:496581  
A/Accession: A00968  
A/Molecule type: Protein  
A/Residues: 106-206, 'S', 208-232, 'A', 234-261, 'N', 263-266, 'S', 266-315, 'N', 317-379 <SMI>  
A/Cross-references: UNIPARC:UPI000002D021  
C/Comment: Secretion of subtilisin is associated with the onset of sporulation, and many  
not necessary for normal sporulation.  
C/Superfamily: Subtilisin, subtilisin homolog  
C/Keywords: extracellular protein; hydrolyase; serine proteinase  
F:1-29/Domain: signal sequence #status predicted <SIG>  
F:30-105/Domain: propeptide #status predicted <APT>  
F:106-379/Product: subtilisin Carlsberg #status experimental <MPR>  
F:128-339/Domain: subtilisin homolog <SBR>  
F:137, 168-325/Active site: Asp, His, Ser #status predicted

|  |                       |              |                    |                 |             |          |
|--|-----------------------|--------------|--------------------|-----------------|-------------|----------|
|  | Query March           | 16.2#;       | Score 554;         | DB 1;           | Length 379; |          |
|  | Best local similarity | 34.1#;       | Pred. No. 3.le-23; |                 |             |          |
|  | Matches .149;         | Conservative | 65;                | Mismatches 149; | Indels 74;  | Gaps 14; |

  

|    |    |   |    |
|----|----|---|----|
| Qy | 3  | GLKALIVILVLGVGSVAAPAEKKVEQVNVNEKNVGLTLPGLFRKIQLKLNPNEISTV | 62 |
|    | :  | :::   | :  |
| Dd | 10 | GMLTAFMLVPFIMARSDSASAAQPAK-----NVEDY-----I                | 41 |

  

|    |    |   |     |
|----|----|---|-----|
| Qy | 63 | IYPENH-REKSIARVLELMGAKRYVVYTHIIIPALADLRVDLVIISLTGGAKLSGYR | 121 |
|    | :  | :::   | :   |
| Dd | 42 | VGFPSGVTASVKDIIKESGSKVDKOFRIINAAKAKDKALKKEVK-----NDPDVA   | 94  |

  

|    |     |  |     |
|----|-----|--|-----|
| Qy | 122 | FIOEDYKYTVSAELEGDESAAQWATVMYLIGYDSGITTIIDPTGIDASHPLDGKVI     | 181 |
|    | :   | :::  | :   |
| Dd | 95  | IYBEDH--VAIALAQTVPYGIPIIKADKYQAQGFAGAVKVAVALDTGIQASHPDL--NVV | 150 |

  

|    |     |  |     |
|----|-----|--|-----|
| Qy | 182 | GMVDPFVNGRSYPYDDHGHTHVASTAAGTGAASNCK--YKMAPCAKLAGIKTVLGADSSG | 239 |
|    | :   | :::  | :   |
| Dd | 151 | GCASFVAGEAAYNTDNGHGHTV---ACTVAALDTTGTGLGAPSVSLYAAYKVLLNSSGSG | 206 |

  

|    |     |   |     |
|----|-----|---|-----|
| Qy | 240 | SISITLIKVEVAVDNKDKYIKVINLSIGSSOSSNGCTDLSQAVNNANDAIGVVCVAAGN | 299 |
|    | :   | :::   | :   |

```

Db      207  TYSGLVSGIEMATN-----GMDVIMSLG---GPSGSATMKAQVDNVAAGVVVVAAGN 259

Qy      300  SGP--NTYVSGPEAAASKYITYGAVDSNDNIASSSSGPTADGRLKEVAVPGVDIIAPR 357
        260  SGGSGGNTNITGYPAKYDSVIAVGAVDNSNRAFPSVQ-----AELEVNAPEGVYSTY 313

Db      358  ASGSTMGFTINNYTRKASGSMATPHVSGVALLQHPSTPDKVTALLETADIYAPK 417
        314  PIST-----YATLNGSTMASPHVAGAAALLILSKHPILNASOVHNRRLSTATYYLG-- 362

Qy      418  EIDIIYVAGRVNYKYA 434
        363  ---SSFTYGGKLINVEEA 377

Db

```

RESULT 5  
139780 subtilisin (EC 3.4.21.62) Sendai precursor - *Bacillus* sp.  
C:Species: *Bacillus* sp.  
C:Date: 19-Jul-1996 #sequence\_revision 19-Jul-1996 #text\_change 05-Oct-2004  
C:Accession: J39780  
R:Yamagata, Y.; Ieshiki, K.; Ichishima, E.  
Enzyme Microb. Technol. 17, 653-663, 1995  
A:Title: Subtilisin Sendai from alkalophilic *Bacillus* sp.: molecular and enzymatic proper  
A:Reference number: J39780; MUID:95529264; PMID:7605625  
A:Accession: J39780  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-382 <RES>  
A:Cross-references: UNIPROT:Q45522; UNIPARC:UP100000B55E9; GB:D29688; NID:G995963; PIDD:R  
C:Genetics:  
A:Gene: aprO  
A:Start codon: TTG  
C:Superfamily: Subtilisin; subtilisin homology  
C:Keywords: hydrolase; serine protease  
/:136-342/Domain: subtilisin homology <SBT>

```

Query Match 15 98; Score 544; DB 2; Length 382;
Best Local Similarity 33.6%; Pred. No. 1,1e-22;
Matches 150; Conservative 69; Mismatches 131; Indels 96; Gaps 17;

Qy      10 VILVLGLVGVSAVAAPBEKK-----VEQVNVNEKNVGLTLPGLFRKIQLNPNE 57
Dp      14 LLLLSLTSATSVAEEQKKQYVIGFENQLQVTEPFVSSDKQ-----55
Qy      58 EISTVTFENHREKEIAVRVLELMGAKRVYVHIIPALADLK---VRDLVLVIGLTGK 114
Dp      56 --SEMSLFAEVNDEISEIMELL-----YEREDIPVSEVELSPEDYKDL-----95
Qy      115 AKLGAVRFIOEDYKYTVASAELE--GLDESAAOVMATYVNNLGYDGGITITGIIDTGIDAS 172
Dp      96 EKDSISITVEIEDIEVTITNQVTPMGI---TRVQAPTAWTRGTYGTVRAVAVLDGTI--ST 150
Qy      173 HPDLQKRTIGWDFPNRGSRYPDDHGHGTHVASIAAGTGAASGK--YCGMAAGAKLGI 230
Dp      151 HPDL--NIRGGSFVPGGSPSYODGNGHGHV---AGTITAAIUNNSIGVGVAAVNAELVAV 204
Qy      231 KVLGADGSGSISTIKGVEMAVDNKDXYGIRKVINLSLSSQSDGTDLSLQAVNNAMADG 290
Dp      205 KVLGANGSGSVSIAQGLQMTAQN---NIHYANISLSGSPV--GSQTEFLAVNQATNAG 257
Qy      291 IVVCVAAGNSGPNPTYVGSPPAASKVITGVGAVDSNDNIASFSSRGPTADGRLEPEVAPG 350
Dp      258 VLVVAATGNNNGSG--TVSGPARYANALAVGATDQNNNRASFSGYGTGLN-----IVA PG 309
Qy      351 VDIAPRASGISMGPIINDYITKASGTSMATTHVSGVGLLIDQAHSGWTPDKYKTKALIER 410
Dp      310 VGI-----QSTYPPGNRYASLSGTSMATPHVAGVAAVLVXQKNPSWSMTQIRQHLTST 360
Qy      411 ADIVAPKEIADIAYAGRVNVYKAIK 436
Dp      361 ATSLGNSN---QFGSGLVNMEAATR 382

```

RESULT 6  
T28159  
Pyrolysin (EC 3.4.-.-) - Pyrococcus furiosus  
C/Species: Pyrococcus furiosus  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T28159  
R/Voorhorst, W.G.B.; Eggen, R.I.L.; Geerling, A.C.M.; Platteeuw, C.; Stenzen, R.J.; Vos, J. Biol. Chem. 271, 20426-20431, 1996  
A/Title: Isolation and characterization of the hyperthermostable serine protease, pyrolysin, from *Pyrococcus furiosus*  
A/Reference number: Z20481; MUID:96355370; PMID:8702780  
A/Accession: T28159  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-1398 <VOO>  
A/Cross-references: UNIPROT:P72186; UNIPARC:UPI000016FD59; EMBL:U55835; NID:G1556462; P1  
A/Experimental source: DSM3638  
C/Genetics:  
A/Gene: pls  
C/Keywords: hydrolase; serine proteinase

Query Match 15.8%; Score 540.5; DB 2; Length 1398;  
Best Local Similarity 22.5%; Pred. No. 9.9e-22;  
Matches 238; Conservative 103; Mismatches 251; Indels 465; Gaps 38;

QY 2 KGLKALLVILVGLV--VGSVAAPKPKVEQVRNVEKNYGLLT---PGLFRKIQKN 54  
DB 4 KGLVLFIALIMLSVVFHFVSAGTPVSSSENSTTILPNOQVTVKEVSOALNALMKQG 63  
QY 55 PNERISIVIFENHREKEIAVRVLEMGAKVRVYHIIPIAIDL---KYRDLVLSGL- 110  
DB 64 PN---MWLIITKEGKLEBAKTELEKIGABITLBNRLMLLYKIKPEKKEINLYSSLE 120  
QY 111 ---TGKAKLSGVRFIOEDYKVTVSABLRLGLDSAAQVM--ATYVMNLGYDSGSGITGII 165  
DB 121 KAMILNREVKLS--PPIYKQVK--TKPELBPKNMNSMTVINALQFIQEFYDSSGVVAVL 178  
QY 166 DTGIDASH---PDIGKVIQWVP---VNGR----- 190  
DB 179 DTGVDNHPPLSLTPDGRKRIEMKQFTDGFVDTSPFSKVVNCTLIINTPQVAGLT 238  
QY 191 ----- 190  
DB 239 LNESTGLMEVYKTVVYVSNVTIGNITSANGIYHFGILPERYFPLFDGQOEDFPYVLVN 298  
QY 191 ----- 194  
DB 299 STANGYDIAYVDTLDYDFTDEVPVLCQYNTVDVAVFSYVYGPVLAIDPNRGYAVF 358  
QY 195 ---DDHGHGTHVASIAGTGAAS-----NGKY-----KGMAPG 224  
DB 359 GMDGHGHGTHVACTVAGYDSNDNDAMDWLSMYSGMEVFSRLYGMWDTYNTTDTVQGVAG 418  
QY 225 AKLAGIKVLGADSGSISTIKGVENAVNKKDYGIKVINLSL--GSSQSGSDGTSLSQAV 283  
DB 419 AQQMAIRVLRSDRGSMMDIIBGMTAA---THGADVLSMGLGNAFLDGTDESVAV 474  
QY 284 NNAAMD--GLVVCVAAAGSGPNTYVSSPAASRVITVAGV----- 323  
DB 475 DELTEKYGVVFAAAGNEPGINIVGSPGATKATIVGAAAPINVGVVYVSGALGPDPY 534  
QY 324 -----SNDNIASPSRGTADGRLKPEVVAPOVDIIAPASGTSKMTPIINDYTKAS 375  
DB 535 GYVYPAATYNNVRIAPFSSSGPRIDGRIKPVVAPGIGY---SSLPMMIGADR---MS 587  
QY 376 GTSMATPHVSGVALLIQA--HPS---WTPDKYKTLIETA-----DIYAPKEIADIAYGA 426  
DB 588 GTSMATPHVSGVALLISGPKEGTYNDIIKKVLESATVLEBDPTGCKTTLDDGH 647  
QY 427 GRVNV-----YKAT-----KYDDYAKL----- 443  
DB 648 GLVAVTKSWEILLKALINGTTLPIVDHMADKSYSDFAEYLGVDVIRGLVARNISIPDIVEMHI 707

QY 444 -----TF-----TGSV----- 449  
DB 708 KYVGDTEYRTFEIYATEPWIKPFVSGSVLENNTEFVLKVDLEGLBGLYVGRITID 767  
QY 450 -----ADKSATH--FDVSGATFYT-----ATLYW 473  
DB 768 PTPVIEDELINTIIVPEKFTPENNNTLTMYDINGEMVTHHFFTVBQVDVLYXMTTWM 827  
QY 474 DTSSSIDLYIDPNC-----NEVDYSYAYVGFKEKVGYNPAGTWTYKVVSYKGAANY 528  
DB 828 DVG-----LYRPDMFVFPYQLDYLPA-----VSNPMPGWELVMTGFPNAPLY 872  
QY 529 QVDVVSQSLSSQSGGNPNPNPNPTPTDTQTFTGSVN--DYMPTSDFTMNV----- 582  
DB 873 E-----SGFLVNIHGVETPS-----VMYINRKYLDITNEFSIEFINITY 913  
QY 583 -----SGATKITGDLTF-----DTSYNDLDLYL 606  
DB 914 APIMATLPIGLGTYNASVESVGDGEFFIKGIEPEGTAEIKIRIGNPSVPSNDDLYL 973  
QY 607 DPNGNLVDRSTSSNSGHEVYANPAPGTWTFVLYAYS 643  
DB 974 DSKGNLVALDGNPFAEEVVEYPRGVYSIVHGS 1010

RESULT 7  
A49778  
high-alkaline serine proteinase (EC 3.4.21.-) precursor - *Bacillus alcalophilus* (strain  
N/Alternate names: subtilisin homolog, high-alkaline  
C/Species: *Bacillus alcalophilus*  
C/Date: 12-Mar-1994 #sequence\_revision 24-Feb-1995 #text\_change 05-Oct-2004  
C/Accession: A49778; J01244  
R/van der Laan, J.C.; Gerltz, G.; Muller, L.J.S.M.; van der Hoek, R.A.C.; Quax, W.J.  
Appl. Environ. Microbiol. 57, 901-909, 1991  
A/Title: Cloning, characterization, and multiple chromosomal integration of a *Bacillus a*  
A/Reference number: A49778; MUID:91282483; PMID:2059048  
A/Accession: A49778  
A/Molecule type: DNA  
A/Residues: 1-380 <VAN>  
A/Cross-references: UNIPROT:P27693; UNIPARC:UPI000020CF5A; GB:D13157; NID:G216231; PIDN:BA02442.1; PID:G  
A/Experimental source: strain P292, ATCC 31408  
A/Note: amino end of mature protein confirmed by peptide sequencing  
Ritakami, H.; Kobayashi, T.; Kobayashi, M.; Yamamoto, M.; Nakamura, S.; Aono, R.; Horikoe  
Biosci. Biotechnol. Biochem. 56, 1455-1460, 1992  
A/Title: Molecular cloning, nucleotide sequence, and expression of the structural gene for  
A/Reference number: J01244; MUID:99043753; PMID:1368952  
A/Accession: J01244  
A/Molecule type: DNA  
A/Residues: 1-195, S', 197-380 <TAK>  
A/Cross-references: UNIPARC:UPI000020CF5A; GB:D13157; NID:G216231; PIDN:BA02442.1; PID:G  
A/Experimental source: *Bacillus alcalophilus* Vedder, ATCC 21522 (*Bacillus sp.* 221)  
C/Keywords: hydrolase; serine proteinase; zymogen  
F.1-27/Domains: signal sequence #status predicted <SIG>  
F.28-111/Domains: activation peptide #status predicted <PRO>  
F.112-380/Product: alkaline serine proteinase #status predicted <MAT>  
F.134-340/Domains: subtilisin homology <SBT>  
F.143,173,326/Active site: Asp, His, Ser #status predicted

Query Match 15.7%; Score 537.5; DB 2; Length 380;  
Best Local Similarity 33.6%; Pred. No. 2.5e-22;  
Matches 147; Conservative 74; Mismatches 140; Indels 77; Gaps 17;

QY 14 LGLVVGSVA-----AAPEKVGQVRNVEKNYGLT---PGLFRKIQKNPNEIS 60  
DB 5 LGKIVASTALLISVAFSSSTASAAEAKKYLIGFNEQ-----EAYSEFVQVANDVVA 59  
QY 61 TVIVFENHREKEIAVRVLEMGAKVRVYHIIPIAIDLKVRDLVYSGLTGKAKLSGV 120  
DB 60 IL-----SEEEVEIETL-----HEFTIPVLSVEISPEP---VDALDELPA-----I 99  
QY 121 RFOEDYKVTVSABLRLGLDSAAQVMATYVMNLGYDSGSGITGIIPTGIDASHPDIQCKY 180

Db 100 STIEDEAVTMA--OSVWGISRVQAPAAHNGLTGSGVAVLDTGI-STHPDL--NI 154

Qy 181 IGWVDFVNRSGYPPDDHGHTHVAISIACTGAASNGK--YKGAAPGAKLAGIKVLGADGS 238

Db 155 RGASAVPPEBPSTQDNGHGTIV---AGTIALNNISIGLVAPAAELIYAVKVLGASGS 210

Qy 239 GSISTTIKVEWAVDNKDKYKIKVNLISGSSQSDGTDLSQAVNNADAGIVCVAAQ 298

Db 211 GSVSSIAQGLEWAGNN---GMHVANLISGSPSPS---ATLEQAVNSATSRGVLVVAASG 263

Qy 299 NSGPNVTYVSGPAASKVITGVAVDSNDNIASFSSGPTADGLKEVVAAPGVDTIAPRA 358

Db 264 NSGAG--SISYPRYANAAVAGTDDNNRPAFSQYAGLD-----IYAPGVNVQSTYP 315

Qy 359 SGTSMGCTPINDYYTKASGTSMATPHYSGVAGALLQHPMTDPKVTALIEIADYAPKE 418

Db 316 GST-----YASINGSMATPHVAGAAALVKKQNPMSNVQITNHLKATATLSGSTN 366

Qy 419 IADIAVAGRVVYKAK 436

Db 367 L-----YSGLVVNAEAATR 380

## RESULT 8

SUBSD  
subtilisin (EC 3.4.21.62) DY - Bacillus subtilis (strain DY)

N:Alternate names: alkaline serine proteinase

C:Species: Bacillus subtilis

A:Variety: strain DY

C:date: 20-Sep-1984 #sequence\_revision 20-Sep-1984 #text\_change 05-Oct-2004

C:Accession: A00969; S02492

R:Nedkov, P.; Oberthur, W.; Bräuntzer, G.

Biol. Chem. Hoppe-Seyler 366, 421-430, 1985

A:title: Determination of the complete amino-acid sequence of subtilisin DY and its comp

A:Reference number: A00969; PMID:85279896; PMID:3927935

A:Accession: A00969

A:Molecule type: protein

A:Residues: 1-274 <MED>

A:Cross-references: UNIPARC:UPI000002D020

A:Experimental source: strain DY

R:Lilova, A.; Kleinschmidt, T.; Nedkov, P.

Biol. Chem. Hoppe-Seyler 368, 1479-1487, 1987

A:title: Reductive alkylation of lysine residues in subtilisin DY.

A:Reference number: S02492; PMID:88134577; PMID:3124865

A:Accession: S02492

A:Molecule type: protein

A:Residues: 1-212-24;27-29;43-45;93-95;123;135-137;140-142;169-171;183-187;221;23

A:Cross-references: UNIPARC:UPI0000055EF8; UNIPARC:UPI0000055EF8; UNIPARC:UPI00001638C;

BFA; UNIPARC:UPI0000172BFB; UNIPARC:UPI0000172BFC; UNIPARC:UPI0000172BFD; UNIPARC:UPI000

C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and many

not necessary for normal sporulation.

C:Superfamily: Subtilisin; subtilisin homology

C:Keywords: extracellular protein; hydrolase; serine proteinase

F:23-234/Domain: subtilisin homology <SBT>

F:32,63,220/Active site: Asp, His, Ser #status predicted

Query Match 15.6%; Score 535; DB 1; Length 274;

Best Local Similarity 44.9%; Pred. No. 2..2e-22;

Matches 132; Conservative 36; Mismatches 90; Indels 36; Gaps 9;

Qy 145 WATYVWNLGYDSGTTIGIDGASHPDLOGKVIYWDVFNHGRSYPYDDHGHTHVA 204

Db 11 IRADKYQAOGYKQANVKGILIDYGAASHDL--KVVGASFSVSGSYNTDNGHGTIV- 67

Qy 205 SIAAGTGAASNGK--YKGAAPGAKLAGIKVLGADSGSISTTIKVEWAVDNKDKYKIKV 262

Db 68 ---AGTVAALDNTTGLVAVPNSVLAIKVLNNSGSGTSAISGLEMATQN---GLDY 120

Qy 263 INSLSSQSDGTDLSQAVNNADAGIYVCAAGNSGP--NTYVSGPAASKVITYG 320

Db 121 INMSLG---GPSSTSLKQAVDKAYASGIVVAAANSGSGSQNTIGYPAKYDVIANG 177

Qy 321 AVDSNDNIASFSSRGPADGRLKEVVAAPGVDTIAPRASGTSMGTPINDYYTKASGTMA 380

Db 178 AVDSNKNRPSFSSVQ-----AELEVMAPGVSVSTYPSNT-----YISLNGTMA 222

Qy 381 TPHVSGVAGALLQHPMTDPKVTALIEIADYAPKEIADIYAGRVVYKAK 434

Db 223 SPHVAGAAALISKPTLSASQVRRRLSTANLQD-----SRYQKGLINVRAA 272

## RESULT 9

JH0778

subtilisin (EC 3.4.21.62) NAT precursor - Bacillus subtilis (strain natto NC2-1)

N:Alternate names: natto proteinase; nattokinase; subtilisin BSP

C:Species: Bacillus subtilis

C:date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 05-Oct-2004

C:Accession: JH0778; J50601; J50517; JC2036

R:Nakamura, T.; Yamagata, Y.; Ichishima, E.

Biosci. Biotechnol. Biochem. 56, 1869-1871, 1992

A:title: Nucleotide sequence of the subtilisin NAT gene, aprN, of Bacillus subtilis (natt

A:Reference number: JH0778; PMID:93113095; PMID:1363081

A:Accession: JH0778

A:Molecule type: DNA

A:Residues: 1-381 <NAK>

A:Cross-references: UNIPROT:P35835; UNIPARC:UPI0000000AB8; GB:D25319; NID:G435439; PIDN:E

R:Sumi, H.

Kagaku To Seibutsu 29, 119-123, 1991

A:title: Natto kinase and fibrinolysis.

A:Reference number: J50601

A:Accession: J50601

A:Molecule type: protein

A:Residues: 107-381 <SUM>

A:Cross-references: UNIPARC:UPI00001565DB

R:Sumi, H.; Nakajima, N.

Nippon Nogelkagaku Kaishi 65, 1125-1127, 1991

A:title: Studies on fibrinolysis enzymes in fermentation food.

A:Reference number: J50517

A:Accession: J50517

A:Molecule type: protein

A:Residues: 107-381 <SU2>

A:Cross-references: UNIPARC:UPI00001565DB

R:Fujita, M.; Nomura, K.; Hong, K.; Ito, Y.; Asada, A.; Nishimuro, S.

Biochem. Biophys. Res. Commun. 197, 1340-1347, 1993

A:title: Purification and characterization of a strong fibrinolytic enzyme (nattokinase)

A:Reference number: JC2036; PMID:94107337; PMID:8280151

A:Accession: JC2036

A:Molecule type: protein

A:Residues: 107-381 <FUB>

A:Cross-references: UNIPARC:UPI00001565DB

C:Genetics:

A:Gene: aprN

A:Start codon: GTG

C:Superfamily: Subtilisin; subtilisin homology

C:Keywords: hydrolase; serine proteinase; zymogen

F:1-29/Domain: signal sequence #status predicted <SIG>

F:30-106/Domain: activation peptide #status predicted <PRO>

F:107-381/Product: subtilisin NAT #status experimental <MAT>

F:129-341/Domain: subtilisin homology <SBT>

F:138,139,170,327/Active site: Asp, Ser, His, Ser #status predicted

Query Match 15.3%; Score 524; DB 2; Length 381;

Best Local Similarity 37.9%; Pred. No. 1.4e-21;

Matches 145; Conservative 53; Mismatches 127; Indels 58; Gaps 15;

Qy 61 TYIVEENREKRIAVRVLGMGAKRYVYHITPAALD---KVPDLVLSGLTGKAKL 117

Db 46 TWSAMSARKKO---VISEKGVQKQPKYVNAATAIDERAVAL-----KQD 91

Qy 118 SGVRFIOEDYKTVTAEL-LEGDESAQVMATYVWNLGYDSGRTIGIDGASHPD 176

Db 92 PSVAAYVEBH---IAHEVAQSVPIGISOIKAPALNSQGYTSNVAVAVIDSIDSHPL 148

Qy 177 QGKVTGWDFVNGRSYPYDD-HGHTHVASIAAGTGAASNGK--YKGAAPGAKLAGIKVL 233

Db 149 --NVRGASFSVPSFTNPYQDGSHTGTHV---AGTIALNNISIGLVAVPNSVLAIVAVKL 202

QY 234 GADGSGISTIIKGVENAVNDKDYKIKVNLISGSSQSDGTDSLQAVNNAMDAGIV 293  
 Db 203 DSGSGGGYVIMINGIEIAIEN----NMDVIMSLG---GPTGSAKTVVDKAVSSGIV 255  
 QY 294 CVAAGNSGP--NTYVGSPPAAAKVITVGVANDNDNIASFSRGPADGLKPEVAPGV 351  
 Db 256 AAAAGNMGSSGSTSTVGVPAKYBSTIIVGVANSSNCRASFSSVGSSEID-----VMAFGV 309  
 QY 352 DIAPRASGSTMGPINDYTTKASGTMATPHVSGVALLQAHPSWTPOKXTALILETA 411  
 Db 310 SIOSTLPFGT-----YGANNGSMATPHVGAALILSKHPTMTNAQVRDLRESTA 360  
 QY 412 DIVAPKEIADIYAGAGRVNYKA 434  
 Db 361 TYLG---NSFYGKGLINVOAA 379  
 RESULT 10  
 JCI085  
 subtilisin (EC 3.4.21.62) precursor - *Bacillus licheniformis*  
 N/Alternate names: alkaline proteinase  
 C/Species: *Bacillus licheniformis*  
 C/Date: 09-Aug-1995 #sequence, revision 19-Oct-1995 #text\_change 05-Oct-2004  
 C/Accession: JCI085  
 R/Lei, H.; Hong, Y.; Zhang, Y.Y.; Shen, T.J.  
 Chinese Biochem. J. 9, 441-447, 1993  
 A/Title: PCR amplifying, cloning and sequencing of the coding sequence of the alkaline F  
 A/Reference number: JCI085  
 A/Accession: JCI085  
 A/Molecule type: DNA  
 A/Residues: 1-275 <LEI>  
 A/Cross-references: UNIPROT:Q9F7C2; UNIPROT:Q53521; UNIPROT:Q9PDP4; UNIPROT:Q9PDP2; UNIF  
 A/Note: The translation of the start codon ATG is not given in this paper  
 C/Superfamily: Subtilisin, subtilisin homology  
 C/Keywords: hydrolase; serine proteinase  
 F/24-235/Domain: subtilisin homology <Str>  
 F/33,64,221/Active site: Asp, His, Ser #status predicted  
 Query Match 15.34; Score 523; DB 2; Length 275;  
 Best Local Similarity 43.5%; Pred. No. 9.9e-22; Indels 36; Gaps 10;  
 Matches 128; Conservative 39; Mismatches 91; Indels 36; Gaps 10;  
 QY 145 VMAVYVNLGYDSGITIGITIDGASHPDLOGKVIQWDFVNGHSYPYDGHGHTVA 204  
 Db 12 IKADKVOAGQFKGAVVAVLDGTIGQASHPD--NVVGGASFYAGAEYNTDGNHGHTV- 68  
 QY 205 SIAGTGAASNGK--YKGAAPGAKLAGIKVILGADGSGSISTIIKGVENAVNDKDKGIV 262  
 Db 69 --AGTVAAALDNTTGVGVSVSLVAVKVLNMSGSGSYSGIYSGIEMATTN----GMDV 121  
 QY 263 INTLSGSSQSDGTDSLQAVNNAMDAGIVGVAAAGNSGP--NTYVGSPPAAAKVITV 320  
 Db 122 INNSLG---GASGTAMKQAVNDIYAGVYVAAAAGNSGSGNTNITIGYPAKIDSIANG 178  
 QY 321 AVDSNDNIASFSSRGPADGLKPEVAPGVADIAPRASGTMGPINDYTTKASGTM 380  
 Db 179 AVDSNSNRASFSSVG-----AELEVAAPG-----AGVSTYPTNTYAT-LNGISMA 223  
 QY 381 TTPVSGVALLQAHPSWTPOKXTALILETADIVAPKEIADIYAGAGRVNYKA 434  
 Db 224 SEHVAGAAALILSKHPTMTNAQVRDLRESTA 360  
 RESULT 11  
 SUBSI  
 subtilisin (EC 3.4.21.62) E precursor - *Bacillus subtilis*  
 N/Alternate names: alkaline proteinase; bacillopeptidase E; extracellular alkaline serin  
 C/Species: *Bacillus subtilis*  
 C/Date: 04-Dec-1986 #sequence, revision 04-Dec-1986 #text change 05-Oct-2004  
 C/Accession: A00972; A26116; I39770; I39778; I39779; S68012; H65586  
 R/Stahl, M.L.; Ferrari, E.  
 J. Bacteriol. 158, 411-418, 1984

A/Title: Replacement of the *Bacillus subtilis* subtilisin structural gene with an in vitro  
 A/Reference number: A00972; MUID:84212198; PMID:6427178  
 A/Accession: A00972  
 A/Molecule type: DNA  
 A/Residues: 1-381 <STA>  
 A/Cross-references: UNIPROT:P04189; UNIPARC:UPI0000142525; GB:K01988; NID:g143519; PIDN:  
 A/Experimental source: strain 1168  
 R/Mong, S.L.; Price, C.W.; Goldfarb, D.S.; Doi, R.H.  
 Proc. Natl. Acad. Sci. U.S.A. 81, 1184-1188, 1984  
 A/Title: The subtilisin E gene of *Bacillus subtilis* is transcribed from a sigma37 promote  
 A/Reference number: A26116; MUID:84144862; PMID:6322190  
 A/Accession: A26116  
 A/Molecule type: DNA  
 A/Residues: 1-155 <MON>  
 A/Cross-references: UNIPARC:UPI000016B997; GB:K01443; NID:g143665; PIDN:AAA22814.1; PID:  
 R/Ikemura, H.; Takagi, H.; Inouye, M.  
 J. Biol. Chem. 262, 7859-7864, 1987  
 A/Title: Requirement of pro-sequence for the production of active subtilisin E in *Escheri*  
 A/Reference number: I39969; MUID:87222417; PMID:3108260  
 A/Accession: I39970  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-156 <IKS>  
 A/Cross-references: UNIPARC:UPI000016B956; GB:M16639; NID:g143521; PIDN:AAA22744.1; PID:  
 R/Henner, D.J.; Ferrari, E.; Perego, M.; Hoch, J.A.  
 J. Bacteriol. 170, 296-300, 1988  
 A/Title: Location of the targets of the hpx-97, sac32(hy), and sacQ36(hy) mutations in  
 A/Reference number: I39778; MUID:8806885; PMID:2447063  
 A/Accession: I39778  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-8 <HEN>  
 A/Cross-references: UNIPARC:UPI000016B7CE; GB:M19125; NID:g142527; PIDN:AAA2245.1; PID:  
 R/Park, S.  
 J. Bacteriol. 171, 2637-2665, 1989  
 A/Title: *Bacillus subtilis* subtilisin gene (apre) is expressed from a sigma-A (sigma-43)  
 A/Reference number: I39779; MUID:89213955; PMID:2496113  
 A/Accession: I39779  
 A/Molecule type: DNA  
 A/Residues: 1-13 <PAR>  
 A/Cross-references: UNIPARC:UPI000016B7CF; GB:M31060; NID:g142529; PIDN:AAA2246.1; PID:  
 A/Experimental source: strain W168, substrain PY79  
 R/Kamal, M.; Hoeseg, J.O.; Kaiser, R.; Shafiqat, J.; Razzaki, T.; Zaidi, Z.H.; Joernvall,  
 FEBS Lett. 374, 363-366, 1995  
 A/Title: Isolation, characterization and structure of subtilisin from a thermostable Bac  
 A/Reference number: S68012; MUID:96069945; PMID:7589571  
 A/Accession: S68012  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 113-323 <KAM>  
 A/Cross-references: UNIPARC:UPI0000172C13  
 R/Kuner, F.; Ogasawara, N.; Mosser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, L.B.; Capuano, V.; Carter, N.M.; Choi  
 A.; Enllich, S.D.; Emmerison, P.T.; Eutelan, K.D.; Birrington, J.; Febre, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A/Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fume, S.; Galizzi, A.; Gallier  
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Hoisappel, S.; Hosono, S.; Hullo, M.F.;  
 Koester, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois,  
 A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,  
 Y.; M.; Ogawa, K.; Ojizawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,  
 A/Authors: Schlöth, S.; Schöfner, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Serox,  
 akuchl, M.; Tamakoshi, A.; Tanaka, T.; Terpersa, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.  
 A/Authors: Yoshikawa, H.F.; Zumeireh, E.; Yoshikawa, H.; Danchin, A.  
 A/Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A/Reference number: A69580; MUID:980404033; PMID:9384377  
 A/Accession: H69586  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-26, 'v', 28-381 <KUN>  
 A/Cross-references: UNIPARC:UPI00006019A; GB:Z99109; GB:AL009126; NID:g2633260; PIDN:CAI  
 A/Experimental source: strain 168





A:Start codon: GTG  
C:Superfamily: Subtilisin; subtilisin homology  
C:Keywords: hydrolase; serine proteinase; zymogen  
F:1-29/Domain: signal sequence #status predicted <SIG>  
F:10-106/Domain: activation peptide #status predicted <ACT>  
F:107-381/Product: subtilisin 3 #status predicted <MAT>  
F:129-341/Domain: subtilisin homology <SBT>  
F:138-170/Active site: Asp, His, Ser #status predicted

Query Match 15.2%; Score 520; DB 2; Length 381;  
Best Local Similarity 37.3%; Pred. No. 2,3e-21;  
Matches 143; Conservative 55; Mismatches 127; Indels 58; Gaps 15;

QY 61 TVIVFENHREKEIAVRVLELMGAKRVRYVHIIPALADL--KYRDLVLSGLTGKAKL 117  
D 46 TWGAMSAAKKKD-----VISEKGVQKQFYVNAATAATLEKAVKEL-----KKD 91  
QY 118 SGVRFIQEDYKVTVAE-LEGDESAQAQVMAVYVNLGYDGGITGITIGDASHPD 176  
D 92 PSVAAYVEEDH---IAHSAQSVPGISQIRAPALHSQGYGSNVKVAVIDSGIDSHPD 148  
QY 177 QGVVIGVNDVNRSPYVD-HGHGTHVASIAAGTGAASNGK--YKGMAPGAKLAGIKVL 233  
D 149 --NVRGASFPSPETNPGQSSSHGTHV---AGTIALNNSIGVLGVSPSASLYAVKVL 202  
QY 234 GADSGSISTIKGVEMAVNDKDKYGIKVINLSIGSSQSDGTDSLQAANNMADAGIVV 233  
D 203 DSGSGQGYWINGIEIAIN---NMDVIMSLG---GPSGSLTKTVVDKAVSSGIVV 255  
QY 294 CVAAGNSGP--NTYTGSPAAASKVITVGAVDNDNIASFSSRGPTADGRLKEVVAAPV 351  
D 256 AAAAGNNGSSSGSTVGYPAKYPSTIIVGAVNSNMQAPSPSSAGSELD-----VMAPGV 309  
QY 352 DIIAPRASGSMGPPIINDYTTKASGTMATPHYSVGALLQAHSPMTPOKVTALILETA 411  
D 310 SIQSTLPFGH-----YGAINGTSMATPHVAGAAALTLKHPVTMAQVDRLESTA 360  
QY 412 DIVAPKEIADIAYGAGRVNYKA 434  
D 361 TYIG---NSFYKXGLINVOAA 379

RESULT 14  
A48373  
high-alkaline serine proteinase (EC 3.4.21.-) precursor - *Bacillus* sp. (strain AH-101)  
N/Alternate names: subtilisin-like thermostable alkaline serine proteinase  
C/Species: *Bacillus* sp.  
C/Date: 03-Feb-1994 #sequence revision 03-Feb-1994 #text change 05-Oct-2004  
C/Accession: A48373; J050714  
R/Takami, H.; Kobayashi, T.; Aono, R.; Horikoshi, K.  
Appl. Microbiol. Biotechnol. 38, 101-108, 1992  
A/Title: Molecular cloning, nucleotide sequence and expression of the structural gene for subtilisin from *Bacillus* sp. strain AH-101  
A/Reference number: A48373; MUID:9308926; PMID:11369007  
A/Accession: A48373  
A/Molecule type: DNA  
A/Residues: 1-361 <TAK>  
A/Cross-references: UNIPARC:UPI0000175C86; GB:S50880; NID:g2e1737; PIDM:AA060421.1; PID:  
A/Experimental source: AH-101  
A/Note: this sequence is inconsistent with the nucleotide translation  
R/Takami, H.; Kobayashi, T.; Aono, R.; Horikoshi, K.  
submitted to JIPID, July 1992  
A/Description: Molecular cloning, nucleotide sequence and expression of the structural gene for subtilisin from *Bacillus* sp. strain AH-101  
A/Reference number: J050714  
A/Accession: J050714  
A/Molecule type: DNA  
A/Residues: 94-334, 'V', 336-361 <TA2>  
A/Cross-references: UNIPARC:UPI0000175C87  
C/Comment: This alkaliphilic *Bacillus* homology to the subtilisins of neutrophilic *Bacillus*  
C/Superfamily: Subtilisin; subtilisin homology  
C/Keywords: extracellular protein; hydrolase; serine proteinase  
F:115-321/Domain: subtilisin homology <SBT>  
F:124,154,307/Active site: Asp, His, Ser #status predicted

Query Match 14.8%; Score 508; DB 2; Length 361;  
Best Local Similarity 33.7%; Pred. No. 9.5e-21;  
Matches 147; Conservative 79; Mismatches 128; Indels 82; Gaps 17;

QY 2 KGIKALIVLVGLVGVSAAPPEKVGQVNVNENYGLTPGLPRKIQKLNPNIEIST 61  
D 3 QSLKLVNLSTVAL-LPMANPAAASEKKEYLLVPE-----DEVA 42  
QY 62 VIIVFENHREKEIAVRVLELMGAKRVRYVHIIPALADLKYRDLVLSGLTGKAKLSGV 121  
D 43 QSVESY-----DVDVHFEFEIPIVHLELBEELKELQNDPNVYA----- 83  
QY 122 FIQEDYKVTVAE-LEGDESAQAQVMAVYVNLGYDGGITGITIGDASHPD 180  
D 84 -IENAEVITISQVPMGISIFISTQ-----QAHNRGIRGNAGARVAVLDGTI-ASHPD--R1 135  
QY 181 IGVNDVNRSPYVDHGHGTHVASIAAGTGAASNGK--YKGMAPGAKLAGIKVLGADSD 238  
D 136 AGGASFISESPSYHDNNGHGHV---AGTIALNNSIGVLGVSPADLYAVVLDNRGS 191  
QY 239 GSISTIKGVEMAVNDKDKYGIKVINLSIGSSQSDGTDSLQAANNMADAGIVVCAAG 298  
D 192 GSILASVAQGIEMALIN---NMHIIIMSLG---STGSSITELAVRANAGILLVGAAG 244  
QY 299 NSGPNYTVGSPAAASKVITVGAVDNDNIASFSSRGPTADGRLKEVVAAPVDTIAPRA 358  
D 245 NTG--RQGVNYPARYGVMAVAADQNGORASFSYGP-----EIEIAPGVNVYS-RY 295  
QY 359 SGRSMGTPINDYTTKASGTMATPHYSVGALLQAHSPMTPOKVTALILETADIAPKE 418  
D 296 TG-----NRRVLSIGSMATPHVAGVAAVLYKSRYSPTNNQIRINQITATYLGSSN 347  
QY 419 IADIAYGAGRVNYKA 434  
D 348 L-----YGNGLVHAGRA 359

RESULT 15  
G83756  
subtilisin-type alkaline proteinase (EC 3.4.21.-) BH0855 precursor [similarity] - *Bacillus*  
C/Species: *Bacillus* halodurans  
C/Date: 01-Dec-2000 #sequence revision 01-Dec-2000 #text change 05-Oct-2004  
C/Accession: G83756  
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirao  
Nucleic Acids Res. 28, 4317-4331, 2000  
A/Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
A/Reference number: A83650; MUID:20512582; PMID:11058132  
A/Accession: G83756  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-361 <STO>  
A/Cross-references: UNIPARC:PA11363; UNIPARC:UPI0000129E77; GB:AP001510; GB:BA000004; NID:  
A/Experimental source: strain C-125  
A/Genes: BH0855  
C/Superfamily: Subtilisin; subtilisin homology  
C/Keywords: hydrolase; serine proteinase  
F:1-25/Domain: signal sequence #status predicted <SIG>

Query Match 14.8%; Score 508; DB 2; Length 361;  
Best Local Similarity 33.4%; Pred. No. 9.5e-21;  
Matches 144; Conservative 80; Mismatches 131; Indels 80; Gaps 15;

QY 2 KGIKALIVLVGLVGVSAAPPEKVGQVNVNENYGLTPGLPRKIQKLNPNIEIST 61  
D 3 QSLKLVNLSTVAL-LPMANPAAASEKKEYLLVPE-----DEVA 42  
QY 62 VIIVFENHREKEIAVRVLELMGAKRVRYVHIIPALADLKYRDLVLSGLTGKAKLSGV 121  
D 43 QSVESY-----DVDVHFEFEIPIVHLELBEELKELKXK-----KDPNVK 82  
QY 122 FIQEDYKVTVAE-LEGDESAQAQVMAVYVNLGYDGGITGITIGDASHPD 181

```

Db      83 ALEKXAEVTIS---QTVPMGISFINTQQAHNNGIFGNGARVAVLDTGI-ASHPDL--R1A 136
QY      182 GWVDEVNGRSYPPYDDHGHGTHVASTAAGTGAASNGK--YKGMAPGAKLAGIKVLGADGSG 239
Db      137 GGASPISESPSYHDNNGHGHV---AGTIALNNSIGVLGVAFPSADLYAVKVLDRNGSG 192
QY      240 SISTIKGVEMAVDNKDKYKIVINLSIGSSQSDQDLSQAVNNAMADAGIVCVAAGN 299
Db      193 SIASVAGIEMALNN---NMHIIINSLG--STGSSSTLELAVNRANNAGIILVGAAGN 245
QY      300 SGPNTYTVGSPAAASKVITVGAVDSDNDNIASFSSRGPTADGRLKPEVVA PGVDIIAPRAS 359
Db      246 TG--RQGVNYPARYSGVMAVAAPDQNGORASFSTYGP-----EIEISAPGVN----- 291
QY      360 GTSMTPIINDYITKASGTSMTAPHVSGVICALILOAHPSTPDKTKTALITETADIAPKEI 419
Db      292 ---NSTYTGNRVSLSGTSMATPHVAGVAAVLVKSRYPSYTNNOIRORINQITATYLGSPSL 348
QY      420 ADIAYAGRVNYYKA 434
Db      349 ----YNGGLVHAGRA 359

```

Search completed: January 6, 2007, 22:32:20  
 Job time : 29 secs

**This Page Blank (uspto)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2007, 22:20:17 ; Search time 192 Seconds  
(without alignments)  
3174.923 Million cell updates/sec

Title: US-10-800-684-5  
Perfect score: 3428  
Sequence: 1 MKGLKALIVILVGLVVGVS.....YAYSTYGMADYQLKAVVYVG 659

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : UniProt 7.2.\*  
1: uniprot\_sprot:\*  
2: uniprot\_trcembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description  |
|------------|--------|-------------|--------|-------|--------------|
| 1          | 3015.5 | 88.0        | 654    | 2     | Q8U0C9 PYRPU |
| 2          | 2705   | 78.9        | 663    | 2     | Q5J1Z5 PYRKO |
| 3          | 980.5  | 28.6        | 561    | 2     | Q8RB02 THETN |
| 4          | 925    | 27.0        | 424    | 2     | Q3C085 THETN |
| 5          | 752    | 21.9        | 1294   | 2     | Q50HM7 STRSH |
| 6          | 748.5  | 21.8        | 1245   | 2     | Q5RL54 STRCO |
| 7          | 737.5  | 21.5        | 795    | 2     | Q5NM24 PARCH |
| 8          | 715.5  | 20.9        | 430    | 2     | Q8ENV1 OCBTH |
| 9          | 703    | 20.5        | 1139   | 2     | Q82139 STRAD |
| 10         | 702    | 20.5        | 1105   | 2     | Q8KKH6 STRVD |
| 11         | 688    | 20.1        | 442    | 2     | Q651P4 BACSD |
| 12         | 688    | 20.1        | 442    | 2     | Q31788 BACIU |
| 13         | 681.5  | 19.9        | 442    | 2     | Q5C1T5 GEOKA |
| 14         | 674.5  | 19.7        | 412    | 2     | Q3C0T3 THETN |
| 15         | 668.5  | 19.5        | 1102   | 2     | P95684 STRAO |
| 16         | 667.5  | 19.5        | 444    | 2     | Q9KXJ7 BACHD |
| 17         | 664.5  | 19.4        | 1220   | 2     | Q9KXJ7 BACHD |
| 18         | 662.5  | 19.3        | 1208   | 2     | Q82B14 STRAM |
| 19         | 658    | 19.2        | 1253   | 2     | Q9P006 STRCO |
| 20         | 655    | 19.1        | 1237   | 2     | Q8GCT4 STRAZ |
| 21         | 644.5  | 18.8        | 412    | 2     | Q8R6R6 THETN |
| 22         | 642.5  | 18.7        | 412    | 2     | Q9AER6 THETO |
| 23         | 635.5  | 18.5        | 369    | 2     | Q3C5Z3 9CIOT |
| 24         | 635.5  | 18.5        | 644    | 2     | Q46C21 METBA |
| 25         | 634.5  | 18.5        | 435    | 2     | Q8EMJ3 OCEIH |
| 26         | 615.5  | 18.0        | 1239   | 2     | Q9P8Z4 STRCO |
| 27         | 564    | 16.5        | 374    | 2     | Q9P941 BACLI |
| 28         | 562    | 16.4        | 379    | 2     | Q6BCN9 BACHO |
| 29         | 562    | 16.4        | 379    | 2     | Q9P941 BACLI |
| 30         | 562    | 16.4        | 379    | 2     | Q651P7 BACIU |
| 31         | 561    | 16.4        | 374    | 2     | Q9F942 BACIU |

|    |       |      |      |   |              |                    |
|----|-------|------|------|---|--------------|--------------------|
| 32 | 560   | 16.3 | 379  | 2 | Q6PNN5 BACLI | Q6PNN5 bacillus 11 |
| 33 | 559   | 16.3 | 374  | 2 | Q9F943 BACLI | Q9F943 bacillus 11 |
| 34 | 559   | 16.3 | 379  | 2 | Q4PKR6 BACLI | Q4PKR6 bacillus 11 |
| 35 | 556   | 16.2 | 379  | 2 | Q53521 BACLI | Q53521 bacillus 11 |
| 36 | 555   | 16.2 | 382  | 1 | SUBT_BACAM   | P00782 bacillus am |
| 37 | 554   | 16.2 | 379  | 1 | SUBT_BACLI   | P00780 bacillus 11 |
| 38 | 551   | 16.1 | 379  | 2 | Q45300 BACLI | Q45300 bacillus 11 |
| 39 | 546.5 | 15.9 | 1358 | 2 | Q8ETM4 OCEIH | Q8ETM4 oceanobacil |
| 40 | 545   | 15.9 | 379  | 2 | Q45301 BACLI | Q45301 bacillus 11 |
| 41 | 544   | 15.9 | 382  | 2 | Q45522 9BACI | Q45522 bacillus sp |
| 42 | 543.5 | 15.9 | 310  | 2 | Q9P8Z3 BACLI | Q9P8Z3 bacillus 11 |
| 43 | 543.5 | 15.9 | 1398 | 2 | Q9P941 PYRMO | Q9P941 pyrococcus  |
| 44 | 542.5 | 15.8 | 567  | 2 | Q48811 COLP3 | Q48811 colwellia p |
| 45 | 541.5 | 15.8 | 1398 | 1 | PUS_PYRPU    | P72186 pyrococcus  |

## ALIGNMENTS

| RESULT 1              | Q8U0C9 PYRPU   | PRELIMINARY | PRT: | 654 AA. |
|-----------------------|--|-------------|------|---------|
| AC                    | Q8U0C9;  |             |      |         |
| DT                    | 01-JUN-2002, integrated into UniProtKB/TrEMBL.                           |             |      |         |
| DT                    | 01-JUN-2002, sequence version 1.   |             |      |         |
| DT                    | 07-FEB-2006, entry version 19.   |             |      |         |
| DE                    | Alkaline serine protease.  |             |      |         |
| GN                    | Ordered locus names: PFI670;   |             |      |         |
| OS                    | Pyrococcus furiosus.   |             |      |         |
| OC                    | Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;    |             |      |         |
| OC                    | Pyrococcus.  |             |      |         |
| OX                    | NCBI_TaxID=261;  |             |      |         |
| RN                    | [1]  |             |      |         |
| RP                    | NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].                           |             |      |         |
| RC                    | STRAIN=Vci / DSM 3638 / ATCC 43587 / JCM 8422.                           |             |      |         |
| RA                    | Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;                            |             |      |         |
| RT                    | "The complete sequence of the Pyrococcus furiosus genome.";              |             |      |         |
| RL                    | Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.                  |             |      |         |
| CC                    | Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  |             |      |         |
| CC                    | Distributed under the Creative Commons Attribution-NonCommercial license |             |      |         |
| DR                    | EMBL; AE010265; AAL81794.1; -; Genomic DNA.                              |             |      |         |
| DR                    | HSSP; Q99405; 1MPT.  |             |      |         |
| DR                    | BioCyc; PPR186497:PFI670-MONOMER; -                                      |             |      |         |
| DR                    | GO; GO:0008233; F:peptidase activity; IEA.                               |             |      |         |
| DR                    | GO; GO:0004289; F:subtilase activity; IEA.                               |             |      |         |
| DR                    | GO; GO:0006508; P:proteolysis; IEA.                                      |             |      |         |
| DR                    | InterPro; IPR007280; Pept arc bac_C.                                     |             |      |         |
| DR                    | InterPro; IPR000209; Pept_s8_S53.  |             |      |         |
| DR                    | Pfam; PF00082; Peptidase_S8; 1.  |             |      |         |
| DR                    | Pfam; PF04151; PFC; 2.   |             |      |         |
| DR                    | PRINTS; PR00723; SUBTILASIN.   |             |      |         |
| DR                    | PROSITE; PS00136; SUBTILASE ASP; UNKNOWN_1.                              |             |      |         |
| DR                    | PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.                              |             |      |         |
| KW                    | PROSITE; PS00138; SUBTILASE_SER; 1.                                      |             |      |         |
| KW                    | Complete proteome.   |             |      |         |
| SQ                    | SEQUENCE 654 AA; 70231 MW; 1CB145A5FE50DB34 CRC64;                       |             |      |         |
| Query Match           | 88.0%; Score 3015.5; DB 2; Length 654;                                   |             |      |         |
| Best Local Similarity | 88.8%; Pred. No. 1.9e-157;   |             |      |         |
| Matches               | 585; Conservative 22; Mismatches 47; Indels 5; Gaps 1;                   |             |      |         |
| QY                    | 1 MKGLKALIVILVGLVVGSAAPAEKKEVQRYNVEKNGYGLTPGLFRKIQKLNPEEIS 60            |             |      |         |
| DB                    | 1 MKGLKALIVILVGLVVGSAAPAEKKEVQRYNVEKNGYGLTPGLFRKIQKLNPEEIS 60            |             |      |         |
| QY                    | 61 TYIVENHREKEIAVRVLEMGAKRYVYHIIIPALADIKVLDLIVISGLTGSKAKLSGV 120         |             |      |         |
| DB                    | 61 TYIVENHREKEIAVRVLEMGAKRYVYHIIIPALADIKVLDLIVISGLTGSKAKLSGV 120         |             |      |         |
| QY                    | 121 REIQDYKVTYSAALEGIDESAAYQWATYVNNLGYDGSGITTIIDTIGIDASHPLQGV 180        |             |      |         |

Db 121 RFIQEDYKTVASAEISGLDSSAQMATTVMNLGYDGSGLTIGITDGDASHEDLQKV 180  
 QY 181 IGWVDFVNGRSYPYDDHGHGTHVASIAAGTGAASNGKYKGMAPGAKIAGIKVLGAADSGS 240  
 Db 181 IGWVDFVNGRSYPYDDHGHGTHVASIAAGTGAASNGKYKGMAPGAKIAGIKVLGAADSGS 240  
 QY 241 ISTTIKVEWAVNDKDKYGIKVINLSLGSQSSDGTDSLQAANNAMDAQIVVCVAAGNS 300  
 Db 241 ISTTIKVEWAVNDKDKYGIKVINLSLGSQSSDGTDSLQAANNAMDAQIVVCVAAGNS 300  
 QY 301 GPNVTYTGSPAAASKVITVGAVDNDNIASFSSSGPTADRLKPEVVAPEGVDIIAPASG 360  
 Db 301 GPNVTYTGSPAAASKVITVGAVDNDNIASFSSSGPTADRLKPEVVAPEGVDIIAPASG 360  
 QY 361 TSMGQPIINDYTTAAGTSMATPHVAGIAALLQHPSTPDYKXIALIETADIVKPEIA 420  
 Db 361 TSMGQPIINDYTTAAGTSMATPHVAGIAALLQHPSTPDYKXIALIETADIVKPEIA 420  
 QY 421 DIAYGAGRVVYRAIKTDYAKLTFTGSVADKGSATTPDVSGATFTATLTYMDTGSSTI 480  
 Db 421 DIAYGAGRVVYRAIKTDYAKLTFTGSVADKGSATTPDVSGATFTATLTYMDTGSSTI 480  
 QY 481 DLVLYDPNGNEVDYATYAYGFEKVGYNPTAGTWTVKVSYKGAANYQVDVYDGSLSQ 540  
 Db 481 DLVLYDPNGNEVDYATYAYGFEKVGYNPTAGTWTVKVSYKGAANYQVDVYDGSLSQ 540  
 QY 541 SGGQNPNDPNPPTPTTDTQFTGTVNDYMDTSDFTNNVNSGATKITGDLTFDTSYND 600  
 Db 541 SGGQNPNDPNPPTPTTDTQFTGTVNDYMDTSDFTNNVNSGATKITGDLTFDTSYND 600  
 QY 601 LDLYLDPNGNVLVDRSTSSNSYEHVEYANPAGTWTFLVYASTYGMADYQLKAVYTG 659  
 Db 601 LDLYLDPNGNVLVDRSTSSNSYEHVEYANPAGTWTFLVYASTYGMADYQLKAVYTG 659  
 QY 659 LDLYLDPNGNVLVDRSTSSNSYEHVEYANPAGTWTFLVYASTYGMADYQLKAVYTG 659  
 Db 659 LDLYLDPNGNVLVDRSTSSNSYEHVEYANPAGTWTFLVYASTYGMADYQLKAVYTG 659  
 RESULT 2  
 OSJ125\_PYRKO  
 ID OSJ125\_PYRKO PRELIMINARY; PRT; 663 AA.  
 AC OSJ125;  
 DT 15-FEB-2005, integrated into UniProtKB/TrEMBL.  
 DT 07-FEB-2006, entry version 1.  
 DE Subtilisin-like serine protease.  
 GN OrderedlocusNames=Fkl689;  
 OS Pyrococcus kodakarensis (Thermococcus kodakarensis).  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae;  
 OC Thermococcus.  
 NCBI\_TaxID=69014;  
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RP STRAIN=KOD1.  
 RA Fukui T., Atomi H., Kanai T., Matsumi R., Fujiwara S., Imanaka T.;  
 RT "Complete genome sequence of the hyperthermophilic archaeon  
 Thermococcus kodakarensis KOD1 and comparison with Pyrococcus  
 genomes."; genome Res. 15:352-363(2005).  
 RL Genome Res. 15:352-363(2005).  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonCommercial License  
 CC EMBL; AP006878; BAD85878.1; -, Genomic DNA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004289; F:subtilisin-like serine protease activity; IEA.  
 DR GO; GO:0006508; P:proteolysis; IEA.  
 DR InterPro; IPR007280; Pept arc bac C.  
 DR InterPro; IPR000209; Pept\_S8\_S53.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR Pfam; PF04511; PFC; 2.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR PROSITE; PS00136; SUBTILASE ASP; UNKNOWN 1.  
 DR PROSITE; PS00137; SUBTILASE HIS; UNKNOWN 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.

KM Complete proteome; Protease.  
 SQ SEQUENCE 663 AA; 70955 MW; 2CE68ACD3888390E CRC64;  
 Query Match 78.9%; Score 2705; DB 2; Length 663;  
 Best Local Similarity 77.7%; Pred. No. 2,3e-140;  
 Matches 517; Conservative 60; Mismatches 80; Indels 8; Gaps 4;  
 QY 1 MKGIALIIVLVILGLVGSVAAAPKKVEQVAVN--EKNGYGLTGTGLPRKIOKLPNBS 58  
 Db 1 MKKFGAVLVALLFLVGLMASSVLAAPKPA--VRNVGQGNVYGLTGLTKKQRMWDDE 58  
 QY 59 ISTVIFENHREKEIAVRVLELMGAKRVYVHTIIPALADLKVRDLVTSGL--TG--GK 114  
 Db 59 VSTIIMFDNQADKEKAVEILDFGAKIKYVHIIPALAVIKIKDILLIAGLMDTGYFGN 118  
 QY 115 AKLSGRFIOEDYKTVASAEISGLDSSAQMATTVMNLGYDGSGLTIGITDGDASHP 174  
 Db 119 AQLSGVQFIQEDYKTVASAEISGLDSSAQMATTVMNLGYDGSGLTIGITDGDASHP 178  
 QY 175 DLQKVIQWVDFVNGRSYPYDDHGHGTHVASIAAGTGAASNGKYKGMAPGAKIAGIKVLG 234  
 Db 179 DLQKVIQWVDFVNGRSYPYDDHGHGTHVASIAAGTGAASNGKYKGMAPGAKIAGIKVLN 238  
 QY 235 ADGSGSISTTIKVEWAVNDKDKYGIKVINLSLGSQSSDGTDSLQAANNAMDAQIVVC 294  
 Db 239 GQSGSISDIINGVMAVQNDKDKYGIKVINLSLGSQSSDGTDSLQAANNAMDAQIVVY 298  
 QY 295 VAAAGNPNTYTGSPAAASKVITVGAVDNDNIASFSSSGPTADRLKPEVVAPEGVDII 354  
 Db 299 VAAAGNPNTYTGSPAAASKVITVGAVDNDNIASFSSSGPTADRLKPEVVAPEGVDII 358  
 QY 355 APASGTSMTPIINDYTTAAGTSMATPHVAGIAALLQHPSTPDYKXIALIETADIV 414  
 Db 359 APASGTSMTPIINDYTTAAGTSMATPHVAGIAALLQHPSTPDYKXIALIETADIV 418  
 QY 415 ARKEIADIVYAGRVVYRAIKTDYAKLTFTGSVADKGSATTPDVSGATFTATLTYMD 474  
 Db 419 KDEIADIVYAGRVVYRAIKTDYAKLTFTGSVADKGSATTPDVSGATFTATLTYMD 478  
 QY 475 TGSSDIDLVLDPNGNEVDYATYAYGFEKVGYNPTAGTWTVKVSYKGAANYQVDVY 534  
 Db 479 NGSDIDLVLDPNGNEVDYATYAYGFEKVGYNPTAGTWTVKVSYKGAANYQVDVY 538  
 QY 535 DGSLSQSGGQNPNDPNPPTPTTDTQFTGTVNDYMDTSDFTNNVNSGATKITGDLTF 594  
 Db 539 DGSLSQSGGQNPNDPNPPTPTTDTQFTGTVNDYMDTSDFTNNVNSGATKITGDLTF 598  
 QY 595 DTSYNDLVLVDPNGNVLVDRSTSSNSYEHVEYANPAGTWTFLVYASTYGMADYQLK 654  
 Db 599 DTSYNDLVLVDPNGNVLVDRSTSSNSYEHVEYANPAGTWTFLVYASTYGMADYQLK 658  
 QY 655 VVYTG 659  
 Db 659 KYYTG 663  
 RESULT 3  
 O8RBJ2\_THETN  
 ID O8RBJ2\_THETN PRELIMINARY; PRT; 561 AA.  
 AC O8RBJ2;  
 DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.  
 DT 01-JUN-2002, sequence version 1.  
 DT 07-FEB-2006, entry version 1.  
 DE Subtilisin-like serine protease.  
 GN Name=Apre2; OrderedlocusNames=TTB0824;  
 OS Thermomacrobacter tengcongensis.  
 OC Bacteria; Firmicutes; Clostridia; Thermomacrobacteriales;  
 OC Thermomacrobacteriaceae; Thermomacrobacter.  
 NCBI\_TaxID=119072;  
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RP STRAIN=MB4 / JCM 11007;  
 RC MEDLINE=21992816; PubMed=11997336; DOI=10.1101/gr.219302;

RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
 RA Tan H., Chen R., Wang J., Yu J., Huang H.;  
 RT "A complete sequence of the *T. tengcongensis* genome.";  
 RL Genome Res. 12:689-700(2002).  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs license  
 CC -----  
 DR EMBL: AE013049; AAM24081.1; -; Genomic\_DNA.  
 DR HSSP: P00782; 2SRT.  
 DR BioCyc: TTN119072:TTB0824-MONOMER; -;  
 DR GO: GO:0082233; F:peptidase activity; IEA.  
 DR GO: GO:0042802; F:protein self binding; IEA.  
 DR GO: GO:0004289; F:subtilase activity; IEA.  
 DR GO: GO:0004086; F:negative regulation of enzyme activity; IEA.  
 DR GO: GO:0006508; F:proteolysis; IEA.  
 DR InterPro: IPR002048; EF\_hand\_Ca\_bd.  
 DR InterPro: IPR007280; Pept\_arc\_bac\_C.  
 DR InterPro: IPR000209; Pept\_s8\_S53.  
 DR InterPro: IPR010259; Prot\_inh\_S8A.  
 DR Pfam: PF00082; Peptidase\_S8; 1.  
 DR Pfam: PF04151; PC; 1.  
 DR Pfam: PF05922; Subtilisin\_N; 1.  
 DR PRINTS: PR00723; SUBTILISIN.  
 DR PROSITE: PS00018; EF\_HAND\_1; UNKNOWN\_1.  
 DR PROSITE: PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE: PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
 DR Complete proteome; Hydrolase; Protease; Serine protease.  
 KW Complete proteome; Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 561 AA; 59696 MW; BA9C5C52F7083A18 CRC64;

Query Match 28.6%; Score 980.5; DB 2; Length 561;

Best Local Similarity 42.3%; Pred. No. 1.3e-45; Matches 246; Conservative 83; Mismatches 174; Indels 79; Gaps 18;

QY 2 KGLKALLIIVLVGLVGVGVAAPPEKVEQVRVEKNYGLTPGLPRKI--QKLNPN-- 56  
 DB 8 KLLSLIALIISLISNELIIVQAPNINLPIDSPKITY----PSLPQKISMDSNKNKLF 63  
 QY 57 -----EISTVIVFENHREKEIAVRVLELWGA-KRVYVHHIIPAIADLKVR 102  
 DB 64 DDLQGLINKPDEEPVILITENKVPDADIFTIAKNIGKFNKHKYKLIIPSIANLTKS 123  
 QY 103 DLLVIVSGLTGAKAKSGVRFIOEDYKVTYSAELE---GLDSAAQVAMTYWVNLGYDS 158  
 DB 124 QINVLV-----KLEIVQIEYDEYVATLDTATKMFGITTAAS-----DFGVTK 168  
 QY 159 GTTIGIDTGDASHPDQ-GKVIGWVDFVNGSRYPYDGHGHTVASTAAGTGAASNGK 217  
 DB 169 NITIIAIIIDGIDGNHVDLSGKAIIGKMDIKNTTPYDNGHGTHTVASTAAGTG-AGNSF 227  
 QY 218 YKGMAPGAKIAGIKVLGADSGSISTTIKGVEMAVDNKQYGIKYNLSLSSGSSDGT 277  
 DB 228 YGVAPDALVGIKVLGADSGSMSTVTAGIDMAVQNKOVYGIKYNLSLSTGTSDDTD 287  
 QY 278 SLSQAVNNAMADGIYVCVAAGNSGPTTVYVSPAAASKITTYGAV---DSNDNIAFS 333  
 DB 288 STSLVNNRAVDSGIYVVVAAGNSGPAKTYIISGPAEKAITVAAMADVLELFPNLASFSS 347  
 QY 334 RGPATDGRLEKPEVAVAGVDIIAPRASGTMGTPINDYTKASGTSMATPHVSGVGLLIQ 393  
 DB 348 RGPATDGRLEKPDIAAPGVNITAAKANS-----VNGYVT-YSGTSMATPFVAGTALMLN 400  
 QY 394 AHPSWTPDKVKTALLETADIAPKEIADIAYGAGRVNYYKAIKYDDYAKLFTGSVAD-- 451  
 DB 401 AHPNLTLPDAKNIIMSTAKSWGPPS-KNVYDYGAGRLDYEAIRVAG---NFRGNNDIVP 455  
 QY 452 -----KGSATHPDVSGATF-VTATLY---WPTGSDIDLTYLDPNGEVNYS 495  
 DB 456 NHYIISGYLPGSRYSPTWTFNANNTSYPIATLILIPDMANYPDFDIYLYDPSGLTIK-S 514  
 QY 496 YTAAYGFEKVGYYNPAGTWTYVVSYKGAANYQVDVSDGS 537

DB 515 STGTORQETITLPSQGTGYVYKVSYNRSGSYVFPDLASGGS 556  
 RESULT 4

ID Q3CJ85 THEBT PRELIMINARY; PRT; 424 AA.  
 AC Q3CJ85;  
 DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.  
 DT 22-NOV-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 4.  
 DE Peptidase 58 and 553, subtilisin, kexin, sedolisin, peptidase, archaeal  
 DE and bacterial C-terminal.  
 GN ORFNames=Tech39DRAFT\_0431;  
 OS Thermoaerobacter ethanolicus ATCC 33223.  
 OC Bacteria; Firmicutes; Clostridia; Thermoaerobacteriales;  
 OC Thermoaerobacteriaceae; Thermoaerobacter.  
 OX NCBI\_TaxID=340099;  
 RN (1)  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=ATCC 33223;  
 RG US DOE Joint Genome Institute (JGI-PGF);  
 RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,  
 RA Hammon N., Israni S., Pittluck S., Richardson P.;  
 RT "Sequencing of the draft genome and assembly of Thermoaerobacter  
 RT ethanolicus 39E.";  
 RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=ATCC 33223;  
 RG US DOE Joint Genome Institute (JGI-ORNL);  
 RA Larimer F., Land M.;  
 RT "Annotation of the draft genome of Thermoaerobacter ethanolicus  
 RT 39E.";  
 RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs license  
 CC -----  
 DR EMBL: AAK001000003; BA065215.1; -; Genomic\_DNA.  
 DR GO: GO:0082233; F:peptidase activity; IEA.  
 DR GO: GO:0004283; F:subtilase activity; IEA.  
 DR GO: GO:0004289; F:proteolysis; IEA.  
 DR GO: GO:0006508; F:proteolysis; IEA.  
 KW Hydrolyase; Protease; Serine protease.  
 SQ SEQUENCE 424 AA; 44126 MW; 335746D47D09B928 CRC64;

Query Match 27.0%; Score 925; DB 2; Length 424;

Best Local Similarity 51.2%; Pred. No. 1e-42; Matches 210; Conservative 52; Mismatches 112; Indels 36; Gaps 11;

QY 155 YDGSITIGIDTGDASHPDQ-GKVIGWVDFVNGSRYPYDGHGHTVASTAAGTGA 213  
 DB 24 YKNDIIVAVITDGTGDSHVDLAGKRVIGMQDFVNGKSPYDNGHGTHTVASTAAGTGG 83  
 QY 214 SNGTKKMAPGAKIAGIKVLGADSGSISTTIKGVEMAVDNKQYGIKYNLSLSSGSS 273  
 DB 84 NN-LYKGVAPGAALVGIKVLDSNGSGTMTVAGIDMAVQNKDVIKIKYNLSLSTGSS 142  
 QY 274 DQDLSQAVNNAMADGIYVCVAAGNSGPTTVYVSPAAASKITTYGAV---DSNDNIA 329  
 DB 143 DQDLSQAVNNAMADGIYVCVAAGNSGPTTVYVSPAAASKITTYGAV---DSNDNIA 329  
 QY 330 SFSRGPATDGRLEKPEVAVAGVDIIAPRASGTMGTPINDYTKASGTSMATPHVSGVGA 389  
 DB 203 SFSRGPATDGRLEKPDIAAPGVNITAAKANS-----INGYVT-YSGTSMATPFVAGTVA 255  
 QY 390 LILQHPSTWTPDKVKTALLETADIAPKEIADIAYGAGRVNYYKAIKYDDYAKLFTGSVAD-- 439  
 DB 256 LMLSANINLAPDAKNIIMSTAKSWGPPS-KNIDYGVGLDYEAIRVAG---NFRGNNDIVP 314

```

Qy 440 ---YAKITFGSVADGSGATHPEDGGAFFATALLY----MDTSGSIDLYLPDMPNNEV 492
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 NHYYAKESLSGSG---RYSDLMTFTNVTDSYSLPAIFILIPDMANYPDPDIYIYLPFTGLV 371
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 493 DYSTVATLYGEKGYGNPFTAGTWTVAVSYKGAANYQVDV-VSDGSLSGS 541
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 372 K-SSTGTGROETIYLPDTGTGYYIKVYSFRSGSNVYPLDLSVGGGSLTLS 420
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

## RESULT 5

|    |  |       |  |
|----|--|-------|--|
| ID | 050HM7   | STRSH | PRELIMINARY; PR1: 1294 AA.   |
| AC | 050HM7   |       |  |
| DT | 07-JUN-2005  |       | Integrated into UniProtKB/TrEMBL.  |
| DT | 07-JUN-2005  |       | Sequence version 1.  |
| DT | 07-FEB-2006  |       | Entry version 4.   |
| DS |  |       | Probable secreted peptidase.   |
| OS |  |       | Streptomyces sphaeroides.  |
| OC |  |       | Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;   |
| OC |  |       | Streptomycineae; Streptomycetaceae; Streptomycetes.  |
| OX | NCBI_TaxID=195949;   |       |  |
| RN |  |       | [1]  |
| RP |  |       | NUCLEOTIDE SEQUENCE.   |
| RC | STRAIN=NCIMB 11891;  |       |  |
| RC | PubMed=15870333; DOI=10.1128/AEM.71.5.2452-2459.2005;                  |       |  |
| RA | Estebanquillo A.S., Gust B., Galm U., Li S.-M., Chater K.F., Heide I.; |       |  |
| RT | "Heterologous Expression of Novobiocin and Chlorobioicin Biosynthetic  |       |  |
| RT | Gene Clusters.";   |       |  |
| RT | Appl. Environ. Microbiol. 71:2452-2459(2005).                          |       |  |
| RN |  |       | [2]  |
| RP |  |       | NUCLEOTIDE SEQUENCE.   |
| RC | STRAIN=NCIMB 11891;  |       |  |
| RA | Steffensky M., Li S.-M., Heide I.;                                     |       |  |
| RL | Submitted (JAN-2003) to the EMBL/Genbank/DBJ databases.                |       |  |
| CC |  |       |  |
| CC |  |       | Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a> |
| CC |  |       | Distributed under the Creative Commons Attribution-NonDerivs license   |
| CC |  |       | -----  |
| DR | EMBL; AY227005; AAP48601.1; -; Genomic DNA.                            |       |  |
| DR | GO; GO:0008233; F:peptidase activity; INA.                             |       |  |
| DR | GO; GO:0004289; F:subtilisin activity; IEA.                            |       |  |
| DR | GO; GO:0006508; P:proteolysis; IEA.                                    |       |  |
| DR | InterPro; IPR003137; PA.   |       |  |
| DR | InterPro; IPR000209; Pept_S8_S53.                                      |       |  |
| DR | Pfam; PF02225; PA; 1.  |       |  |
| DR | Pfam; PF00082; Peptidase_S8; 1.  |       |  |
| DR | PRINTS; PR00723; SUBTILISIN.   |       |  |
| DR | PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.                            |       |  |
| DR | PROSITE; PS00138; SUBTILASE_SER; 1.                                    |       |  |
| DR | SEQUENCE 1294 AA; 135576 MW; 804C7P9A0DDER896 C6C64;                   |       |  |

```

Query Match 21.9%; Score 752; DB 2; length 1294;
Best Local Similarity 45.1%; Pred. No. 1.3e-32;
Matches 173; Conservative 62; Mismatches 125; Indels 24; Gaps 9

QY 62 VVFNHREKEIAVLVEIM--GAKRYVYVHIIPATAADLKVDL-----LVISGLTG 112
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||::
DB 133 LTVSSGKKAEATVRRITRALDDGSAVRRL-ESIDARAIVSPADLGAFWKQLAPRGPEI 191
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

QY 113 GKAKLSGRFIOEDYKVTVSABLBSGLDESAQVMATYVNLGYDGSGITTIITDGIDAS 172
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
DB 192 NSAALDAPKWLDDGRVAS-----LPRSTQIAPVPMWSGIRGERVKVAVLDTGADQT 246
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

QY 173 HPDLQOKVIGWDFPNRGSYPYDPDHGCHTHVASITAGTGAASNGKYKGMAGKLAGITV 232
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
DB 241 HPDLAGRIAAADF-SSSSGTADRGCHTHVASIVSGSGASGGRGVAAPAEELMTGCV 305
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

QY 233 LGADSGSISITTIKGVEMAVDNCKDKYGIKVINLSGSSOSDGTDSLSCQVNN-AWDAGI 291
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
DB 306 LGDDGFGSPSYIAGMEWAALK-----GAEVNNMISGSDAPSDGTDPMISLAVNELSSSSGA 361
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

QY 292 VVCVAAGNSGPRITTVGSPAAASKVITYTGAVDSNDNIASFSSGCP-TADGRLEKPVAVAG 350
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

```

|    |     |   |     |
|----|-----|---|-----|
| Db | 362 | LFVAAAGNSGPGSGTIGSGAADALTVGAARDSDSLAEFSRSPRSGDAEVRKDPVAPG | 421 |
| Qy | 351 | VDIIAPRASGTMGPIINDYYTKASGTSMATPHVSGVALILLOHPMTDKXYVALIET  | 410 |
| Db | 422 | VGIYAARCTGTGMSPVDVDDGTAAAGTSMATPHVGAALLAQNHPRMNAQLKDALVST | 481 |
| Qy | 411 | ADIVAPKEIADIALYGAQRVNTYKA                                 | 434 |
| Db | 482 | ARTIAGQVTE--QGGRIDILAA                                    | 503 |

## RESULT 6

|  |                            |
|--|----------------------------|
| 09KRL54_SIRCO  | ID                         |
| 09RL54_SIRCO   | PRELIMINARY; PRF; 1245 AA. |
| 09RL54   | AC                         |
| 01-MAY-2000, integrated into UniProtKB/TrEMBL.                         | DT                         |
| 01-MAY-2000, sequence version 1.                                       | DT                         |
| 07-FEB-2006, entry version 21.   | DT                         |
| Probable secreted peptidase.   | DE                         |
| Ordered locusNames=SC000432; ORNNames=SCF51A.10;                       | GN                         |
| Streptomyces coelicolor.   | OS                         |
| Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;           | OC                         |
| Streptomycinae; Streptomycetaceae; Streptomyces.                       | OC                         |
| NB01_TaxID=1902;   | OX                         |
| [1]  | RP                         |
| NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].                         | RP                         |
| STRAIN=A3(2) / M145;   | RC                         |
| MDLID=21596410; PubMed=12000953; DOI=10.1038/41714a;                   | RC                         |
| Bentley S.D., Chater K.P., Cerdeno-Parraga A.-M., Challis G.L.,        | RA                         |
| Thompson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H.,        | RA                         |
| Harper D., Bateman A., Brown S., Chanda G., Chen C.W., Collins M.,     | RA                         |
| Cronin A., Fraser A., Goble A., Hiddle J., Hornsby T., Howarth S.,     | RA                         |
| Huang C.-H., Kleiser T., Larke L., Murphy L.D., Oliver K., O'Neill S., | RA                         |
| Rabbinowitch E., Rajandream M.A., Rutherford K.M., Ruter S.,           | RA                         |
| Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,   | RA                         |
| Marrion T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,   | RA                         |
| Woodward D.A.;   | RA                         |
| "Complete genome sequence of the model actinomycete Streptomyces       | RT                         |
| coelicolor A3(2).";  | RT                         |
| Nature 417:141-147(2002).  | TL                         |

```
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonCommercial 4.0 International license
CC -----
DR EMBL, AL939105; CAB6662.1; -; Genomic_DNA.
DR HSSP, P00782; ISUE.
DR BioCyc; SCOE1902:SCOD432-MONOMER; -.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilisin-like serine protease activity; IEA.
DR GO; GO:0006508; P:proteolysis; IEA.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010221; VCSB.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; Subtilisin.
DR TRIGRAMS; TIRG01965; VCSB_repeat; 1.
DR PROSITE; PS00137; SUBTILASIN_HTS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Complete proteome.
SQ SEQUENCE 1245 AA; 130896 MW; 74EE92DB9CAID60 CRC64;
```



```

QY      177  OQKVI GWNVFVNRSXPYDHDHGHTFVASIAAGTGAANGKXKMAIPAKIAGIKVLGAD 236
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      252  AGRVAAAKDP -SSSSGTNDVFGHGHVVASIVGSSGAHSGSSRGVAPAPARILVCKVLGDD 310
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      237  GSGSIIITIKGVEMAVDNKDKYGIKIVINLSLSSQSOSDGTSLSOAVNN -AMDAGIVCV 295
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      311  GFSEBEOVLAGHEMMDQ----GADVNMNLSGSSGATGTCTPMQALNDLSRRTGTLFVY 366
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      296  AAGNSG -PNTTYVGSPPAAASKVITGVAVDSNDNIASFSSRGP -TADGRLKEVVAPGVDI 353
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      367  AAGNEGEQGPRTVGSFGADALITGAVDRDLSLAFSSRGPRLGDADVXDPVATPVGVI 426
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      354  IAPRAGTSKNGPIINDYTTKASGTSMATPHVSGVALLLOHPSTPDKXTALILEADI 413
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      427  VAPARAASGAMGPVBEBHYTAAGTSMATPHVAGAAALLAOQHPPMTGQALDIALISTAVT 486
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      414  VAPKEIADIAYAGAGRNVYKAIKYDDYAKLFTGSVADKGSATHTPDSGATFTATLWY 473
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      487  VDGQKYTE--QGGGRIDVRAA---GLGAVTATGTL-----VMGPFTSRDTEPVTSKRY 535
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      474  DTGSSDIDLVL 484
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      536  -TNSSDDEVTL 545
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 7

ID OSNM24\_9ARCH PRELIMINARY; PRT; 795 AA.  
AC OSNM24;  
DT 04-JAN-2005, integrated into UniProtKB/TrEMBL.  
DT 04-JAN-2005, sequence version 1.  
DT 07-FEB-2006, entry version 6.  
DE Alkaline serine protease.  
DE ORFNames=orf17;  
OS uncultured archaeon.  
OC Archaea; environmental samples.  
NCBI\_TaxID=115547;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=16329940; DOI=10.1016/j.femsec.2004.12.004;  
RA Ertel C., Kemnitz D., Kube M., Rieke P., Chin K.-J., Dedys S.,  
RA Reinhardt R., Conrad R., Liesack W.,  
RT Retrieval of first genome data for rice cluster I mechanisms by a  
RT combination of cultivation and molecular techniques.";  
LT COMB Microbiol. Ecol. 53:187-204(2005).  
LL

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/term>

CC Distributed under the Creative Commons Attribution-NonCommercial License

CC

DR EMBL, CR636856, CAHQ4782.1, -, Genomic DNA.

DR GO: 00088233, F:peptidase activity, IEA.

DR GO: 0042802, F:protein self binding, IEA.

DR GO: 0004289, F:subtilase activity, IEA.

DR GO: 0004306, P:negative regulation of enzyme activity, IEA.

DR GO: 0006508, P:proteolysis, IEA.

DR InterPro: IPR011964, Beta\_ptc\_yvhn.

DR InterPro: IPR000209, Pept\_G6\_S53.

DR InterPro: IPR010259, Pept\_tmh\_S8A.

DR Pfam: PF00082, Peptidase\_S5, 1.

DR Pfam: PF05922, Subtilisin\_N, 1.

DR PRINTS: PR00723, SUBTILISIN.

DR TIGRFAMs: TIGR02276, beta\_ptc\_yvhn, 3.

DR PROSITE: PS00136, SUBTILASE ASP, 1.

DR PROSITE: PS00137, SUBTILASE HIS, UNKNOWN 1.

DR PROSITE: PS00138, SUBTILASE\_SER, UNKNOWN\_1.

DR PROSITE.

DR SEQUENCE. 795 AA; 82723 MW; 629536E8B3D16BB8 CRC64;

Q

|                           |       |                   |             |             |
|---------------------------|-------|-------------------|-------------|-------------|
| Query Match               | 21.5% | Score 737.5;      | DB 2;       | Length 795; |
| Best Local Similarly      | 32.3% | Pred.No. 4.6e-32; |             |             |
| Matches 200; Conservative | 90;   | Mismatches 208;   | Indels 121; | Gaps 17;    |

|    |     |   |     |
|----|-----|---|-----|
| Qy | 51  | OKNPBEEISTYIVENHREKIIARVELMLGAKVRYVYHIIPIADLKRPDLVSGL               | 110 |
| Dd | 26  | KTIIDNQKGASLLSQGAASKAKRDLVYAGGKVKTKINIIDVADLDPDEVAL---              | 82  |
| Qy | 111 | TGGKATLSGVRFIOEDYKTVVSAALEGLDESAAQMAATYVNL--GYDSSGITIGIIDTG         | 168 |
| Dd | 83  | ---KAR-PDVEHSVBD-----SIAYADLDEVYHNVAASVAMSTVPGYTGKGVNVVSIDSG        | 133 |
| Qy | 169 | IDASHFDLOGKITYGNVDFPNGRSGSYYYDDHGHGTHVASIAGTGAASNGKTKGMAPGAULA      | 228 |
| Dd | 134 | IDATHFDLAGKITYLWKDFLNDRAATPYDDPGHGTFAAGIASSGAMSGEYKGLAYDASLFF       | 193 |
| Qy | 229 | GIATFLADSGSISTIIKGVEMAVNKKDKYGIKTYINLSIGSSQSDGTDLSQAVNNAWD          | 288 |
| Dd | 194 | GVYVLPNSTGTAAYSDIIAALDKSVGNH---ADVISMSLSNPTH---IQALDAVHNAAE         | 246 |
| Qy | 289 | AGIIVCVAAAGNSGPNITYVVGSPAAASKITYTGAVDSDNUNIASFSSRGPTFADGRKEPVNA     | 348 |
| Dd | 247 | NGVYVVCASAGNGPKYKSIKPCGDSPDVIAAGSVMSBRLSSFSRSRPTDGRKIPDIVA          | 306 |
| Qy | 349 | PGVDIIAPASGTSNGTPIINDYTTKASTGNMATPVSGVALIIQAHSPMPDKVKALI            | 408 |
| Dd | 307 | VEGYTVSYRSGSTTGNPLGQYTYCYASGISAACPOVSAASAILIQANGLSTPEITKOVLI        | 366 |
| Qy | 409 | ETADIVAPKEIADI-----AYGAGRVNVYKAIAKYDDYKALTFTGVSADKGSATHTFDVSGA      | 464 |
| Dd | 367 | RMT-----YHLSDBTYPENQGGKGRINIAPAL--NEVLQVPTPTPTPTPATPTPTPTPT         | 419 |
| Qy | 465 | TFPTATLTYLWDTGSSDIDLXLVDPNGENVDYSTATYGEBEKGYNNPNA--GTYTAKVVS        | 521 |
| Dd | 420 | ATPTAT-----PATPTPTPTPTPTPT-----PATPTPTPTPTPTPT                      | 440 |
| Qy | 522 | YKGAANYQVDVVSDDLSQGGGNPNPNPNPTPT-----TDTOTFT                        | 564 |
| Dd | 441 | -----TVPTPTATPT | 485 |
| Qy | 565 | GSVNDYWDTSITFTWN-----VNSGATKITGD-LTFDTSYNDLXLVYDPNGNLVDR            | 615 |
| Dd | 486 | PFPTPTPPSKTFVNSYNSGNTFVSDIGTGNITVTGTIITGSRPY-----IVHSPDGSIV--       | 539 |
| Qy | 616 | STSSNSYEHVEYANPARGT   | 634 |
| Dd | 540 | YVAVEGLNRVALIISPPANT  | 558 |

## RESULT 8

| ID                | GENE1   | OCE1H                            | PRELIMINARY | PRT | 430 AA |
|-------------------|---|----------------------------------|-------------|-----|--------|
| Q8ENV1            | OCE1H   |                                  |             |     |        |
| AC                | OCE1H   |                                  |             |     |        |
| DT                | 01-MAR-2003   | integrated into UniProtKB/TrEMBL |             |     |        |
| DT                | 01-MAR-2003   | sequence version 1               |             |     |        |
| DT                | 07-FEB-2006   | entry version 19                 |             |     |        |
| DE                | Intracellular alkaline serine proteinase.                     |                                  |             |     |        |
| CN                | OrderedAccession=O82375                                       |                                  |             |     |        |
| OS                | Oceanobacillus theymansii                                     |                                  |             |     |        |
| OC                | Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus |                                  |             |     |        |
| NCBI_TaxID=182710 |   |                                  |             |     |        |

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;  
RX MEDLINE=22220767; PubMed=12235376; DOI=10.1093/nar/gkx526;  
RA Takami H., Takaki Y., Uchiyama I.;  
RT "genome sequence of *Yersinia enterocolitica* bioserotype 4/O:3 isolated from the Iheyama  
RT Ridge and its unexpected adaptive capabilities to extreme  
RT environments";  
RL Nucleic Acids Res. 30:3927-3935(2002).

CC  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC  
DR EMBL; BA000028; BAC14331.1; -; Genomic\_DNA.  
DR HSSP; Q99405; IMPT.  
DR BioCyc; O1HE182710:OB2375-MONOMER; -.  
OR

DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004289; F:subtilase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis; IEA.  
 DR InterPro; IPR000209; Pept\_S8\_S53.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR PROSITE; PS00136; SUBTILASE ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE SER; 1.  
 DR Complete proteome; Hydrolase; Protease; Serine protease.  
 DR SEQUENCE 430 AA; 45839 MW; 6D09A99BCE1E310F CRC64;  
 SQ  
 Query Match 20.9%; Score 715.5; DB 2; Length 430;  
 Best Local Similarity 46.8%; Pred. No. 3.4e-31;  
 Matches 148; Conservative 56; Mismatches 95; Indels 19; Gaps 6;  
 QY 138 LDESAQVMATYWNIGYDSSGITIGITDGDASHDLOGKIVGWDFNGSYPPDH 197  
 DB 120 LPTASSINADVKEGSLTQGGSTIAVIDGTH-PHEHLEGRIIGFADFYKQTEPYDDN 178  
 QY 198 GHGTHVASTAGGAANGKYGKAPGAKLAGIKVLGADSSGSIPTIKGVMAVDNKK 257  
 DB 179 GHGTHVAGDAGAGALSDGYQGPAPANLVGVKLNKTSGLSTVIGIDMCIONSK 238  
 QY 258 YGKIVNLSIGS--SQSSDGTDSLQAVNNAMPAGIVCAAGNSGPTVTGSPAAASK 315  
 DB 239 YNNINISLSIGSATPEABG-DEYVNAVETAMNGVVCVAAGNSGPDKTGSPGISPK 297  
 QY 316 VITVGAVDN-----DNIAFSRSRGPADGRLKEBVAAPGVDTIAPRASG-----TSM 363  
 DB 298 VITVGAADDNNTAERSDSDSAEFSRSRPTIDGLTKPMLTPGVDTVSLRAPSFDIKTNK 357  
 QY 364 GTINYYTKASGTSMATPHVSGVGLILQHPMTDPKXTALLETADIVAKETADIA 423  
 DB 358 SARVGSNTYLSLSGTSMATPICAGIVAQLOSLSLTPNQVKEKMEACQDLGQ---SPNV 414  
 QY 424 YGAGRVVYKAIKYDD 439  
 DB 415 QGAGVLAANLININE 430  
 RESULT 9  
 Q82139 STRAW PRELIMINARY; PRT; 1139 AA.  
 ID Q82139 STRAW  
 AC Q82139  
 DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.  
 DT 01-JUN-2003, sequence version 1.  
 DT 07-FEB-2006, entry version 16.  
 DE Putative subtilisin-like protease.  
 DE Ordered accession names=SAV3319;  
 GN Streptomyces avermitilis.  
 OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=33903;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
 RX MEDLINE=22608306; PubMed=12692562; DOI=10.1038/nbt820;  
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
 RA Sakaki Y., Hattori M., Omura S.;  
 RA "Complete genome sequence and comparative analysis of the industrial  
 RT microorganism Streptomyces avermitilis";  
 RT Nat. Biotechnol. 21:526-531(2003).  
 RN [2]  
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
 RX MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;  
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;  
 RA "Genome sequence of an industrial microorganism Streptomyces  
 RT avermitilis: deducing the ability of producing secondary  
 RT metabolites";

RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonCommercial License  
 CC  
 CC EMBL; BA000030; BAC71030.1; -, Genomic\_DNA.  
 DR HSSP; Q99405; 1MPT.  
 DR MEROPS; S08.069; -.  
 DR BioCyc; SAV3319-MONOMER; -.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004289; F:subtilase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis; IEA.  
 DR InterPro; IPR002860; Glyco\_Hydro\_BNR.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR PROSITE; PS00136; SUBTILASE ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE HIS; UNKNOWN\_1.  
 DR PROSITE; PS00138; SUBTILASE SER; 1.  
 DR Complete proteome; Protease.  
 DR SEQUENCE 1139 AA; 118193 MW; EC3BD234FA94FA0 CRC64;  
 SQ  
 Query Match 20.5%; Score 703; DB 2; Length 1139;  
 Best Local Similarity 34.3%; Pred. No. 5.6e-30;  
 Matches 205; Conservative 80; Mismatches 215; Indels 98; Gaps 22;  
 QY 110 LNCGRKASGVRFIDQYKVTYSALEGLDESAQVMATYWNIGYDSSGITIGITDGI 169  
 DB 210 VTNMGDTYASGVAVHWLD-----GTRKASLDKSVQCGALPTMAAGDGGKIAVLDTGV 264  
 QY 170 DASHPDLQKIGVWDFVNGRSYPPDDHGHGTHVASTAGTAAGSKYKGAAPGAKLAG 229  
 DB 265 DATHPDLKQVAKESKIF-SAAADADHPGHGTHVASTAGTAKSKYKGAAPGAKLAG 323  
 QY 230 IKVLGADSSGSIPTIKGVMAVDNKKYGIKVINLSLSQSSDGTDSLQAVNN-AMD 288  
 DB 324 GKVLDDTSGSDSGILAGMEMAAEQ-----GADVNNLSLGGGDTPE-IDPLEAEVNLSEB 378  
 QY 289 AGIVCAANGSP-NTYTVGSPAAASKVITGAVDSDNNINAFSRSRGPADGRLKEPV 347  
 DB 379 KGLFAIAAGNEEFGEQITGSPSAADALTVGAVDSDKLASFSSRGGLDGLKIPDVT 438  
 QY 348 APGVDTIAPRASGTSGTPIIN--DYTKASGTSMATPHVSGVGLILQHPMTDPKXY 404  
 DB 439 APGVDTIAPRASVTDQEVGQKPDYLTISGTSMATPHVAGAAALTKQHPMWSAEIK 498  
 QY 405 TALIETADIVAPKEADIAVYAGRVNVYAIKYDYAKLFTTGSVADKGSAT-----H 457  
 DB 499 GALTGSA---KGGKYTPFOGSGRIADVRAIK-----OSVINPNVSFGLIQWPH 546  
 QY 458 TPDVSATVTYNTL-YMDGSSDIDLXL-----YDPNGNEVDYTYVYGEKVGYNPTA 512  
 DB 547 TDDKP---VTQQLYRNLSGTSVTLNLASTATNPRG-----VAAPSGFFKLGATKTV 556  
 QY 513 GTWTVKVSYKGAANYQDVVSDSGLS-----SGGAGNP-----N 547  
 DB 597 PAGKASVDF--IVNKKLGTTDGAISAVYTAGGGQYRTAAVQREVSYDVLKHD 654  
 QY 548 PNPENPPTPTDTQTFGVSVDVY---DTSDTFTNVSAGATKLTGDLTFD---TSYN 599  
 DB 655 RDKGPVNVSTDLTGVSGLAADKMFAPYDASGTVKRVKGNFLINASFADDEDFTKGA 714  
 QY 600 DLDLYYDNGNLVDST---SSNSAEHYEYANPAAGTMTFLVYAYSTGMADYOLKA 654  
 DB 715 D---WIAQPLSVTKNTTVVDARKAKPVDITVPDKG-----AKSAPSPDITVEA 762  
 RESULT 10  
 Q8KXK6 STRVD PRELIMINARY; PRT; 1105 AA.  
 ID Q8KXK6 STRVD  
 AC Q8KXK6  
 DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.

DT 01-OCT-2002, sequence version 1.  
 DT 07-FEB-2006, entry version 16.  
 DE 1,4-dihydropyridine enantioselective esterase precursor.  
 GN Name=dnpA,  
 OS Streptomyces viridosporus.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomyces; Streptomyces; Streptomyces.  
 NCBI\_TaxId=67581;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=A-914;  
 RX MEDLINE=22034940; PubMed=12039725;  
 RX DOI=10.1128/AEM.68.6.2716-2725.2002;  
 RA Aritaawa A., Matsufuji M., Nakashima T., Dobashi K.,  
 RA Yoshikawa T., Yamada S., Momose H., Taguchi S.;  
 RT "Streptomyces serine protease (DHP-A) as a new biocatalyst capable of  
 RT forming chiral intermediates of 1,4-dihydropyridine calcium  
 RT antagonists.";  
 RL Appl. Environ. Microbiol. 68:2716-2725(2002).  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs License  
 CC EMBL; AB007809; BAC00500.1; -; Genomic\_DNA.  
 DR HSSP; P00782; 2SST.  
 DR MEROPS; S08.069; -;  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004289; F:subtilase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis; IEA.  
 DR InterPro; IPR000209; Pept\_S8\_S53.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PRO00723; SUBTILISIN.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 DR HydroLase; Protease; Serine protease; Signal.  
 FT SIGNAL 1 204 Potential.  
 SQ SEQUENCE 1105 AA; 114126 MW; 6AB34DB8B70579AB CRC64;  
 Query Match 20.5%; Score 702; DB 2; Length 1105;  
 Best Local Similarity 33.1%; Pred. No. 6.2e-30;  
 Matches 205; Conservative 63; Mismatches 227; Indels 124; Gaps 22;  
 QY 110 LFGKAKLGSVRIQDQDKYTVSALEGLDESAQVMATYVMVLGYDGGSTTIGIIDTG 169  
 DB 182 YTMGDTFAGIAHWLD-----GVRRALDFTSVGOIGAPKAWAGYDGKGVKLAVIDTG 236  
 QY 170 DASHPLQGVKVGWVDFVNGRSYFYDDHGHVASTAAGTGAASNGKYKMAPAKLAG 229  
 DB 237 DTSHPLKGRVTKAKFTAPG-AGDKVGHGHVASTAAGTGAASNGKYKMAPAKLAG 295  
 QY 230 IKVLGADSGSISTIIKGVEMAVDNKDKYKIVINLSGSSGSSDGTSLQAANN-AMD 288  
 DB 296 GKVLDSGFGDSDGLAGHMAA-----AQADAVNNSLGMDTPR-TDPLEAAVDKLSAE 350  
 QY 289 AGIVVVAAGNSPNTYTVGSPAAASKVTVGAVDSNDNIASSSSGPR-TADRLKPEVY 347  
 DB 351 KGVLFIAAGNEPPE--SIGSPSADALVGVAVDDKDLADSSFGPRGDAIKRVDY 408  
 QY 348 APGVNDIAPRASGTMGPIND---YYTKASGTMATPHYSGVGAIILOHPSWTPDKYK 404  
 DB 409 APGVNDITTAASAEENDIGQEVGEPAGYMTISGTSMAITPHVGAALILKQHPMTSAELK 468  
 QY 405 TALIEFADIVAPRIADIAGAGRVNVAIK----- 436  
 DB 469 GAL--TGSTKGGK-YTPEFGSGSRIGADKALQOTVIADPVSVSGVQVQWPHTDPEVTK 525  
 QY 437 -----YDYAKLFTGTVADKSGATHTFDVSGATFYATLYMPDGGSDIDLYLDPN 488  
 DB 526 LITRNIGTQDVTIKLSTATDPKGAAPAGFTLIGATTVVPA---GGASAVMTADTRL 562  
 QY 489 GNEVDYSYTAAYYGFKEKGYVNPAGTWTVCVSY-----KGAANYQVDVVS DGS 537

DB 583 GGTVDGAYSAYVATGGQVTRTAAAVQREVESYDVTRHIGRDKPTEHLTDLIGYAG 642  
 QY 538 LSQSGGNGNPNPNPPTTDTOTFTGSVNDY-----W-----D 572  
 DB 643 LSGKRGY-----ABATDTATLRPKGYTLVDSWIAKDPGTLKGDMLVQPKLSV 693  
 QY 573 TSDTFPMVNSGATKITGDLTF-DTSYNDLDE---LYDPNGNLVDRSTSSNSYEHVEYA 628  
 DB 694 TKDT-TLTLDAATTK-AADITVPDPKAPLSATIGTYVTAG--IGIGASMSFADVRWA 749  
 QY 629 NPAP-----GTWT 636  
 DB 750 HIGPEAPGIRQTMNGQWT 768  
 RESULT 11  
 ID Q65IP4\_BACLD PRELIMINARY; PRT; 442 AA.  
 AC Q65IP4; Q62U51;  
 DT 25-OCT-2004, integrated into UniProtKB/TrEMBL.  
 DT 25-OCT-2004, sequence version 1.  
 DT 07-FEB-2006, entry version 15.  
 DE AprX (Alkaline serine protease).  
 GN Name=aprx; OrderedLocustNames=BL00903, BL102185;  
 OS Bacillus licheniformis (strain DSM 13 / ATCC 14580).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 NCBI\_TaxId=279010;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RX PubMed=15337318; DOI=10.1155/000079829;  
 RA Veith B., Herzberg C., Steckel S., Feesche J., Maurer K.H.,  
 RA Ehrenreich P., Baumer S., Henne A., Liesegang H., Mekl R.,  
 RA Ehrenreich A., Gottschalk G.;  
 RT "The complete genome sequence of Bacillus licheniformis DSM13, an  
 RT organism with great industrial potential.";  
 RL J. Mol. Microbiol. Biotechnol. 7:204-211(2004).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RX PubMed=15461803; DOI=10.1186/gb-2004-5-10-r77;  
 RA Key M.W., Kamalya P., Nelson B.A., Brody-Karlin S.D., Zaretsky E.J.,  
 RA Tang M.W., Lopez de Leon A., Xiang H., Gusti V., Clausen I.G.,  
 RA Olsen P.B., Raemussen M.D., Andersen J.T., Joergensen P.L.,  
 RA Larsen T.S., Sorokin A., Bolotin A., Lapidus A., Galleron N.,  
 RA Ehrlich S.D., Berka R.M.;  
 RT "Complete genome sequence of the industrial bacterium Bacillus  
 RT licheniformis and comparisons with closely related Bacillus species.";  
 RL Genome Biol. 5:RESEARCH077.1-RESEARCH077.12(2004).  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs License  
 CC EMBL; AB017333; AAU41070.1; -; Genomic\_DNA.  
 DR EMBL; CP000002; AAU23708.1; -; Genomic\_DNA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004289; F:subtilase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis; IEA.  
 DR InterPro; IPR000209; Pept\_S8\_S53.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PRO00723; SUBTILISIN.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; UNKNOWN\_1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 DR Complete proteome; Protease.  
 KW SEQUENCE 442 AA; 48227 MW; 834B9CC27DFD83A CRC64;  
 Query Match 20.1%; Score 688; DB 2; Length 442;  
 Best Local Similarity 37.9%; Pred. No. 1.2e-29;  
 Matches 188; Conservative 86; Mismatches 135; Indels 38; Gaps 11;  
 QY 45 GLPKIKQNPBEISTVVPENHREKEIAVAVLELMA-----KRYVYHIIPALADLK 100  
 DB 41 GFPEKI-LMKKKIPLVIEFERKDCQHKGYOLVNEMIGSKRNKIKHRSNVS CCSAE-- 96

QY 101 VRDLIVISGLTGKRAKLSGVRFIQEDYVTVSAELBGLDESAQAQVMTYVNLGYDGGSI 160  
 Db 97 ----VTPSSIQSLISECGDRIKRYLNRKYKALLVVAVSSSHAKEVNR---NQTLTGKV 149  
 QY 161 TIGIIDTIDGASHPDLQGVIGWDFVNGRSPYVDHGHGTHVSIAGGAANGCYKG 220  
 Db 150 TVAVITDGV-YPHEDLERIRAFQDFIQRTEPDDHGHGCHGADGAGNAASSGGYRG 208  
 QY 221 MAPGAKLGIIVLGAADGSGSISTTIKGYEMAV---DNKQRYGKIVINLSIGSS---QSS 273  
 Db 209 PAPEAEIVGVKVLKMGSGSLETVYQGVDMCIQFNKENPDD-PDIIISMGLAALAYEN 267  
 QY 274 DGTSLGQAVNANMADAGIVCVAAAGSGPNNTYTGSPAAASKVITVGAVD-----SND 326  
 Db 268 EEDPVPVAAVAAWADAGIVCVAAAGNSGPDQITIASPVSCKITVVALDRTVRSRDD 327  
 QY 327 NIASFSSRGPTADRLKPEVAPGVDIAPASGTSMG-----TPIDYTKASGTSMT 381  
 Db 328 DVASYSSRGPTIYQGVVDDLVPEGTNITSLRSPSGFLDKQTKRVGTKTMTLSGISMAT 387  
 QY 382 PHSVGVALIIQAHPSWTPDKVKTALLETADIYAPKEIADIYAGAGRVNYKAIKYD 438  
 Db 388 PICAGIALIILQAPGTEPDEVKQLMDGTDLM--KDRPNVYGAGYINAEQSVPOD 442

## RESULT 12

031788\_BACSU PRELIMINARY; PRT; 442 AA.  
 ID 031788\_BACSU PRELIMINARY; PRT; 442 AA.  
 AC 031788;  
 DT 01-JAN-1998, integrated into UniProtKB/TrEMBL.  
 DT 01-JAN-1998, sequence version 1.  
 DT 07-FEB-2006, entry version 31.  
 DE Alkaline serine protease.  
 GN Name=apix; OrderedlocusNames=BSU17260;  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 NX NCBI\_TaxID=1423;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377; DOI=10.1038/36786;  
 RA Kunst F., Ogawara N., Moser I., Albertini A.M., Alloni G.,  
 Azevedo V., Besserer M.G., Bessieres P., Bolotin A., Borchert S.,  
 Bortiss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 Brouiller S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 Choi S.-K., Codani J.-J., Conneron I.F., Cummings N.J., Daniel R.A.,  
 Denicot F., Devine K.M., Duesterhoeft A., Ehrlich S.D., Emerson P.T.,  
 Denton K.-D., Ertington J., Fabret C., Ferrari E., Foulger D.,  
 Fritz C., Glaser P., Goffeau A., Golightly E.U., Grandi G.,  
 Grim S.-Y., Glaser P., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 Giuseppe G., Guy B.J., Hega K., Halech J., Harwood C.R., Henaut A.,  
 Hilbert H., Holsappel S., Hosono S., Hullo M.-F., Itaya M.,  
 Jones L.-M., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M.,  
 Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S.,  
 Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J.,  
 Lazarevic V., Lee S.-M., Levine A., Liu H., Masuda S., Mauel C.,  
 Medigue C., Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S.,  
 Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B.,  
 Park S.-H., Patro V., Pohl T.M., Portetelle D., Porwollik S.,  
 Prescott A.M., Pressac B., Pujic P., Purnelle B., Rapoport G.,  
 Rey M., Reynolds S., Rieger M., Rivolta C., Rocha S., Roche B.,  
 Rose M., Sadate Y., Sato T., Scanlan R., Schleich S., Schroeter R.,  
 Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P.,  
 Shin B.-S., Solido B., Sorokin A., Taconi E., Takagi T., Takahashi H.,  
 Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P.,  
 Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vanler H.,  
 Vasseroiti A., Viari A., Wambuit R., Wedler B., Wedler H.,  
 Weitzengger T., Winters P., Wipat A., Yamamoto H., Yaneane K.,  
 Yasumoto K., Yata K., Yoshida K., Yoshikawa H.-F., Zumbstein E.,  
 Yoshikawa H., Zanchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis."  
 RL Nature 390:249-256(1997).

CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonCommercial License  
 CC  
 DR EMBL: Z99113, CAB13610.1, -, Genomic DNA.  
 DR PIR: A69587, A69587.  
 DR HSP: Q99405, 1MP1.  
 DR BioCyc: BSUB1423:BSU1727-MONOMER; -.  
 DR GO: GO:0008233, F:peptidase activity, IEA.  
 DR GO: GO:0004289, F:subtilase activity, IEA.  
 DR GO: GO:0006508, F:proteolysis, IEA.  
 DR InterPro: IPR00209, Pept\_S8\_S53.  
 DR Pfam: PF00882, Peptidase\_S8, 1.  
 DR PRINTS: PR00723, SUBTILISIN.  
 DR PROSITE: PS00136, SUBTILASE\_ASP, 1.  
 DR PROSITE: PS00137, SUBTILASE\_HIS, 1.  
 DR PROSITE: PS00138, SUBTILASE\_SER, 1.  
 KW Complete proteome; Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 442 AA; 47906 MW; B996446AB87BADF2 CRC64;

Query Match 20.1%; Score 688; DB 2; Length 442;  
 Best Local Similarity 39.7%; Pred. NO. 1.2e-29;  
 Matches 167; Conservative 78; Mismatches 132; Indels 44; Gaps 14;

QY 43 TPELFKRI--QKLNPEEISTVIVEENHREK--EIAVRVLEL-MGAKRVYVYHIIIPAIA 96  
 Db 35 TPCFLHKFPETKLNRRKMSVIEFEBCGHEGFGMAGSVLQKRSKLSKFNKNCSS 94  
 QY 97 ADLKTRDLIVISGLTGKRAKLSGVRFIQEDYVTVSAELBGLDESAQAQVMTYVNLGYDGGSI 154  
 Db 95 AR-----VTPSALHSLISECSNIR-----KVIYNRVVKLLDPATASAKKEVVRNGQ 142  
 QY 155 -YDGSGITIGITIDTIGDASHPDLQGVIGWDFVNGRSPYVDHGHGTHVSIAGGAANGCYKG 213  
 Db 143 TLTGKGVYAAVVDITL-YPHEDLERIRAFQDFIQRTEPDDHGHGCHGADGAGNAASSGGYRG 201  
 QY 214 SNGTKKGAAPGAKLGIIVLGAADGSGSISTTIKGYEMAV---DNKQRYGKIVINLSIGSS 269  
 Db 202 SSGQYRGPAPPEANLIGVKNKQSGTGLADIIEGVEMCIQYNEPDE-PIDIMSMSSIGG 260  
 QY 270 S---QSSGCTDLSQAVNANMADAGIVCVAAAGSGPNNTYTGSPAAASKVITVGAVDEN 325  
 Db 261 DALRYDHEQDEPLVAVEAMASAGIVCVAAAGNSGPDQITIASPVSCKITVVALDRTVRSRDD 320  
 QY 326 -----DNIAFSSRGPTADRLKPEVAPGVDIAPASGTSMG-----TPIDYTKASGTSMT 374  
 Db 321 TASSDDIVASFSSRGPTIYQGVVDDLVPEGTNITSLRSPSGFLDKQTKRVGTKTMTLSGISMAT 380  
 QY 375 SGTSMATPHVSGVALIIQAHPSWTPDKVKTALLETADIYAPKEIADIYAGAGRVNYKAIKYD 434  
 Db 381 SGTSMATPICAGIALIILQAMPDLTPDEVKELKNGTD--KMKDBDPVYGAIVNAENS 438  
 QY 435 I 435  
 Db 439 V 439

## RESULT 13

051315\_GEOKA PRELIMINARY; PRT; 442 AA.  
 ID 051315\_GEOKA PRELIMINARY; PRT; 442 AA.  
 AC 051315;  
 DT 01-FEB-2005, integrated into UniProtKB/TrEMBL.  
 DT 01-FEB-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 10.  
 DE Intracellular alkaline serine protease.  
 GN OrderedlocusNames=GK0210;  
 OS Geobacillus kaustophilus.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.  
 NX NCBI\_TaxID=1462;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=HTA426;  
 RX PubMed=15576355; DOI=10.1093/nar/gkh970;

**R**A Takami H., Tokaki Y., Chee G.-J., Nishi S., Shitamamura S., Suzuki H.,  
**RA** Matsui S., Uchiyama I.;  
RT "Thermoadaptation trait revealed by the genome sequence of  
RT thermophilic Geobacillus kaustophilus.",  
RL Nucleic Acids Res. 32:6292-6303(2004).  
CC -----  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs license  
CC -----  
DR EMBL; BA000043; BADY4495.1; - Genomic DNA.  
DR GO; GO:0008233; F.peptidase activity; IEA.  
DR GO; GO:0042802; F.protein self binding; IEA.  
DR GO; GO:0004289; F.subtilase activity; IEA.  
DR GO; GO:0043086; P.negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P.proteolysis; IEA.  
DR InterPro; IPRO00209; Pept\_58\_S53.  
DR InterPro; IPRO10259; Prot\_inh\_S8A.  
DR Pfam; PF00982; Peptidease\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; SUBTILISTN.  
DR PROSITE; PS00136; SUBTILASE\_ASP; UNKNOWN\_1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
KM Complete proteome; Hydrolyase; Protease; Serine protease.  
SQ SEQUENCE 442 AA; 48629 MW; 30F9F60ADD09D20 CRC64.

Query Match                  19.9%; Score 681.5; DB 2; Length 442;  
Best Local Similarity        41.9%; Pred. No. 2.6e-29;  
Matches 167; Conservative 65; Mismatches 132; Indels 35; Gaps 12;

DQ 59 ISTVIVERNREKE-IAVRVLEMGKRRYYTHIIIPALADLRDLVIIGLTGGRKL 117  
::|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
DB 53 LSVLIIEVENAGMEVAVAKAERDHFRMKRHHFRHVFYS--RYVPALAEOLHPKYK- 109  
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
DY 118 SGVAFIEDYKVTVSAELEGDESA-AOVMTVVWNLYDSGGITGIIDPIDASHPDL 176  
-----KVEFKRIYKALLINNAVPSANAKRVAANGEL--SGKVITIADVGTI-YPHEDL 160  
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
DY 177 QGVKTGMWDFPNRGSYPRYDDHGHTHVASIAGTGAA-SNGKKYGMAPAKLAGIKVLGD 236  
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
DB 161 EGRIAAPFDVFNGLTTTPYDDNHGHCHCGDAAGNRMSDGLAYGRAYEAMLLGYVKLDRS 220  
::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
DY 237 GGSGISTTIKCWEANAVDNNKXY---GIKVINLSLGSSGS--DGTSLSQAANNMWDG 290  
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
DY 221 GSGLTEIMRMGEICMDYNENHPSKRDIILISLSIGEPPPIENDPDLVOYAEGAMEQG 280  
::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
DY 291 IVVCVAAAGNSGPNTTYTGPSPAASKVTIVGAVD-----SNONIAFSFGPTADRQL 342  
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
DB 281 IWVCVAAAAGNSBPNGTLISSPGISDTLVGALLDHDTRAFTTRADDVVASFPSSGPREYGT 340  
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
DY 343 KPEVVAACGVDIR--APPASGSTSMG--TPINDYYTKASTMATPVHVGVALILIQAHS 397  
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
DB 341 KPDLIVECVNIILRAPRSPFLDKNMKGSRVEDHYISMSTGMTATPICGIYALMLQAKFN 400  
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
DY 398 WTPDKVKTALIETADIYAPKEIIDAIYAGAGRNVYKAIK 436  
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
DB 401 ATPDEIRKALKDGADLMKGRD--PWVYAGAGYGNCKRAIE 437  
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

RESULT 14  
O3CCT3 THEBT PRELIMINARY; PRT; 412 AA.  
AC O3CCT3;  
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.  
DT 22-NOV-2005, sequenced version 1.  
DT 07-FEB-2006, entry version 4.  
DE Peptidase S8 and S53, subtilisin, kexin, sedoimin.  
GN ORFNames=Teich39DPRAFT 1862;  
OS Thermococcuslactitolan ATCC 33223.  
OC Bacteria; Firmicutes; Clostridia; Thermococcaceae;  
OC Thermococcuslactitolan; Thermococcuslactitolan.  
NCBI\_TaxID=340099;  
LN [1]

RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=ATCC 33223;  
 RA US DOE Joint Genome Institute (JGI-RGF);  
 RG Copeland A., Lucas S., Lipidus A., Barry K., Dettler J.C., Glavina T.,  
 RA Hammond N., Israni S., Pritchard S., Richardson P.,  
 RT "Sequencing of the draft genome and assembly of Thermoaerobacter  
 RT ethanolicus 39E.";  
 RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=ATCC 33223;  
 RG US DOE Joint Genome Institute (JGI-ORNL);  
 RA Larimer F., Land M.;  
 RT "Annotation of the draft genome of Thermoaerobacter ethanolicus  
 RT 39E.";  
 RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs License  
 CC CC  
 DR EMBL; AA001000019; EAO64328.1; -; Genomic DNA.  
 DR GO: GO:0008233; F:peptidase activity; IEA.  
 DR GO: GO:0004289; F:subtilase activity; IEA.  
 DR GO: GO:0006508; P:proteolysis; IEA.  
 KM Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 412 AA; 44421 MW; CEEBHF082716EF8 CRC64;  
  
 Query Match 19.7%; Score 674.5; DB 2; Length 412;  
 Best Local Similarity 40.1%; Pred. No. 5.9e-29;  
 Matches 157; Conservative 66; Mismatches 130; Indels 39; Gaps 11;  
  
 QY 63 IFEENREKEIARVLEIMAKRYVYHIIIPAIALD---KVRDLVTSGLTGKATLGG 119  
 DB 40 VILXSHSYITLKEKTEHKGGSIKFLPITIKASVNLPCDKLNF-----ATLKG 89  
  
 QY 120 VREFIOEDKYVTASAELEGIDESAAQVMATYVN-IGYDSGGITGIIDTGIDASHDL- 176  
 DB 90 IHRFAEDSAVKQLTYI-----ATQELASRNDLDGYTGKGVITAFIDTGI-YRHPFTK 142  
  
 QY 177 -QGKVLGWDVFNNGRSFPHYDDHGHVVASIAGTGAASNGKTKGMAPGAKLAGIYVLGA 235  
 DB 143 PKRRIAFHNVVNGKQPYDNDHGTHVADAGNGVANSNGKKGVAPEANIYAAVKVLD 202  
  
 QY 236 DSGSISTIIKGVEMVNDKDKGIGIVNLISLSSOSSDG-TDLSQAVNMADAGIYVC 294  
 DB 203 YRGGLSSDIIAGQMILNDKDKNIRVSLISGETSLPFLDPLRVGVDTLWRKGLVVT 262  
  
 QY 295 VAAAGSGPNTYTVGSPAAASKVITTVGAVD-----SNDNIASFSSRGPTADGRLEPV 347  
 DB 263 VAAAGSGPNTNTITSPGTSKNAITTVGAVDDKRTSDISDBEIAQFSGRG--SPLYKRDIV 320  
  
 QY 348 APEVDIIAPASGTSNG--TPINDYTTKASGTSMAATPHVSGVICALILQAHBSPWPKVK 404  
 DB 321 APEGVKIVTASENVVPEGADEVITINKKRYRTVGTGSMATPMAAGAALLLEKNPMLTINVQIK 380  
  
 QY 405 TALIETADIVAPKEIADIAGAGRVNVYKIK 436  
 DB 381 NIKKSTN-IKIDDAGLMTGSGMNIIEALK 410

CC Streptomyces, Streptomyces, Streptomyces.  
 OX NCBI\_TaxID=1887;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=S-1253;  
 RX MEDLINE=97144528; PubMed=8990295;  
 RA Suzuki M., Taguchi S., Yamada S., Kojima S., Momose H.;  
 RT "A novel member of the subtilisin-like protease family from  
 Streptomyces albobacillus";  
 RL J. Bacteriol. 179:430-438(1997).  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs License  
 CC -----  
 DR EMBL: D83672; BAI12040.1; -; Genomic\_DNA.  
 DR HSSP: P00782; 2SBT.  
 DR MEROPS: S08\_069; -.  
 DR GO: GO:0008233; F:peptidase activity; IEA.  
 DR GO: GO:0004289; F:subtilase activity; IEA.  
 DR GO: GO:0006508; P:proteolysis; IEA.  
 DR InterPro: IPR002860; Glyco\_hydro\_BNR.  
 DR InterPro: IPR00209; Pept\_S8\_S53.  
 DR Pfam: PF00082; Peptidase\_S8; 1.  
 DR PRINTS: PR00723; SUBTILISTIN.  
 DR PROSITE: PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE: PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
 KM Protease.  
 SQ SEQUENCE 1102 AA; 114128 MW; F9E4AD2590FE559E CRC64;

Query Match 19.5%; Score 668.5; DB 2; Length 1102;  
 Best Local Similarity 32.2%; Pred. No. 4.3e-28;  
 Matches 202; Conservative 71; Mismatches 228; Indels 127; Gaps 22;

QY 109 GLTGKAKSGVRFQEDYVTVSAELGLDESAQVMATYWNLTGDSGITIGITIDTG 168  
 DB 172 GVTDGRTASGVARWLD---GVRKASLDTSVGOIGTPKAWAGYDGKGVKIAVLDTG 226  
 QY 169 IDASHPDLQGVKIGWVDFVNGRSYFYDDHGHVHVASIAGTGAASNGKYGMAPGA 228  
 DB 227 VDAIHPDLKQGVYASKNFTSAPT-TGVVGHGHVHVASIAGTGAOSKGYKVAPOAKII 285  
 QY 229 GIVYLADSGSGSTIIKGYEMAVDNKKYGIKVINLSGSSQSDGTSLSQAVNN-AM 287  
 DB 286 NGKVLADAGFGDSGIIAGMEMAA---AQADIVNMSLGMDTPE-TDPLEAAVDTLSA 340  
 QY 288 DAGIVCVAAGSGPNYTVGSPAAASKVITGAVDSNDNIAFSRGP-TADGRLKEVY 346  
 DB 341 EKGIIFRAIAGNRPQ--STGSPASASALTGAVDDKDLADFSSTGPRIGDAVAPDL 398  
 QY 347 VAPGVDIAPRASGTSGMTPIINDY---YTKASGTSMATPHVSGVALLIQAHPSWT 403  
 DB 399 TAPGVDTIITASAGNDIAKEVGEKPAQYMTISGTMATPHVGAALLIKQHPWKYAEI 458  
 QY 404 KTLA-IETADIVPKIADIAYAGRVNYKAIY-----DD--Y 440  
 DB 459 KQALITASTKD---GKTPPEGSGRVQVDKAITQTIVIAPVSLSPFVQVQPHADKPV 514  
 QY 441 AKLTF---TGSVADKGSATHT-----FDVSGATFVTATLYMDGSSDIDLXYLD 486  
 DB 515 KULTYRNLTGEDVTLLKLTSTATGPKGAAPAGFTLIGASTLVPA---NGTASVDV 571  
 QY 487 PNGNEVDYSYATAYGFEKVGYNPTAGTWTVKVSY-----KGAANYQ----- 529  
 DB 572 RLGGAVDGTYSAYVATGAGSVRTAAVERREVESYNTVLKVLDRSGKATANYMAYLSGL 631  
 QY 530 -----VDVVSQSLG---OSGG-----GNPNPNPNPT 555  
 DB 632 TGIGKDRSYAPYEDAGAVRPKGGYVLDASVYLGADPETWGADWLQPKLDVTRNTT 691  
 QY 556 PTTDTQ-----TFTGSVNDVWDTSDTFPMNVNSGATKITGDLTFDTISYNDL 608

DB 692 VTVDAKRAKPVKTVPGKAKAKQFASADYTIETNDSAVSGMWLENYSGFRSAHL----- 746  
 QY 609 NGNLVDRSTSSNSYEHVEYANPAPGTWT 636  
 DB 747 -GPQITNGTILSQOW-NTHFSNGAKAQYT 772

Search completed: January 6, 2007, 22:31:22  
 Job time : 197 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2007 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2007, 22:32:37 ; Search time 123.5 Seconds  
(without alignments)  
2471.729 Million cell updates/sec

Title: US-10-800-684-5

Sequence: 1 MKGLKALIVILVIGLVVGS.....YAVSTYGNADYQKAVVYVG 659

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_Main:\*  
1: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
3: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
4: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match Length | ID   | Description                              |
|------------|--------|--------------------|------|--|
| 1          | 3428   | 100.0              | 659  | US-09-841-553-5<br>Sequence 5, Appl1     |
| 2          | 3428   | 100.0              | 659  | US-10-800-684-5<br>Sequence 5, Appl1     |
| 3          | 2997.5 | 87.4               | 654  | US-09-841-553-35<br>Sequence 35, Appl1   |
| 4          | 2997.5 | 87.4               | 654  | US-10-090-624-16<br>Sequence 16, Appl1   |
| 5          | 2997.5 | 87.4               | 654  | US-10-800-684-35<br>Sequence 35, Appl1   |
| 6          | 2997.5 | 87.4               | 654  | US-10-888-588-16<br>Sequence 16, Appl1   |
| 7          | 2914   | 85.0               | 659  | US-09-841-553-1<br>Sequence 1, Appl1     |
| 8          | 2914   | 85.0               | 659  | US-10-090-624-12<br>Sequence 12, Appl1   |
| 9          | 2914   | 85.0               | 659  | US-10-800-684-1<br>Sequence 1, Appl1     |
| 10         | 2914   | 85.0               | 659  | US-10-888-588-12<br>Sequence 12, Appl1   |
| 11         | 2349.5 | 68.5               | 522  | US-09-841-553-3<br>Sequence 3, Appl1     |
| 12         | 2349.5 | 68.5               | 522  | US-10-090-624-4<br>Sequence 4, Appl1     |
| 13         | 2349.5 | 68.5               | 522  | US-10-800-684-3<br>Sequence 3, Appl1     |
| 14         | 2349.5 | 68.5               | 522  | US-10-888-588-4<br>Sequence 4, Appl1     |
| 15         | 1918   | 56.0               | 412  | US-10-090-624-1<br>Sequence 1, Appl1     |
| 16         | 1918   | 56.0               | 412  | US-10-888-588-1<br>Sequence 1, Appl1     |
| 17         | 703    | 20.5               | 1139 | US-10-156-761-10856<br>Sequence 10856, A |
| 18         | 662.5  | 19.3               | 1079 | US-10-112-488-39<br>Sequence 39, Appl1   |
| 19         | 662.5  | 19.3               | 1079 | US-10-673-860-8<br>Sequence 8, Appl1     |
| 20         | 662.5  | 19.3               | 1208 | US-10-156-761-13251<br>Sequence 13251, A |
| 21         | 655    | 19.1               | 1237 | US-10-314-657-4<br>Sequence 4, Appl1     |
| 22         | 655    | 19.1               | 1237 | US-10-473-193-4<br>Sequence 4, Appl1     |
| 23         | 562    | 16.4               | 379  | US-09-813-408-11<br>Sequence 11, Appl1   |
| 24         | 561    | 16.4               | 372  | US-09-813-408-9<br>Sequence 9, Appl1     |
| 25         | 560    | 16.3               | 379  | US-11-156-062-23<br>Sequence 23, Appl1   |
| 26         | 557    | 16.2               | 379  | US-09-920-118-14<br>Sequence 14, Appl1   |
| 27         | 557    | 16.2               | 379  | US-10-202-339-2<br>Sequence 2, Appl1     |

|    |       |      |     |   |                   |                    |
|----|-------|------|-----|---|-------------------|--------------------|
| 28 | 555   | 16.2 | 382 | 4 | US-10-090-624-31  | Sequence 31, Appl1 |
| 29 | 555   | 16.2 | 382 | 4 | US-10-104-693-2   | Sequence 2, Appl1  |
| 30 | 555   | 16.2 | 382 | 5 | US-10-800-684-45  | Sequence 45, Appl1 |
| 31 | 555   | 16.2 | 382 | 5 | US-10-888-588-31  | Sequence 31, Appl1 |
| 32 | 555   | 16.2 | 382 | 5 | US-10-498-714A-2  | Sequence 2, Appl1  |
| 33 | 555   | 16.2 | 382 | 6 | US-11-127-727-2   | Sequence 2, Appl1  |
| 34 | 555   | 16.2 | 382 | 6 | US-11-127-947-2   | Sequence 2, Appl1  |
| 35 | 554   | 16.2 | 379 | 4 | US-09-813-408-10  | Sequence 10, Appl1 |
| 36 | 554   | 16.2 | 379 | 4 | US-10-146-905A-10 | Sequence 10, Appl1 |
| 37 | 554   | 16.2 | 379 | 5 | US-10-503-706-1   | Sequence 1, Appl1  |
| 38 | 553   | 16.1 | 380 | 4 | US-10-146-905A-12 | Sequence 12, Appl1 |
| 39 | 551   | 16.1 | 379 | 3 | US-09-813-408-13  | Sequence 13, Appl1 |
| 40 | 551   | 16.1 | 382 | 4 | US-10-146-905A-8  | Sequence 8, Appl1  |
| 41 | 550   | 16.0 | 380 | 3 | US-09-813-408-19  | Sequence 19, Appl1 |
| 42 | 549.5 | 16.0 | 378 | 3 | US-09-813-408-14  | Sequence 14, Appl1 |
| 43 | 548.5 | 16.0 | 380 | 5 | US-10-836-959-2   | Sequence 2, Appl1  |
| 44 | 548.5 | 16.0 | 380 | 5 | US-10-476-463-2   | Sequence 2, Appl1  |
| 45 | 547   | 16.0 | 382 | 4 | US-10-033-325-2   | Sequence 2, Appl1  |

## ALIGNMENTS

RESULT 1  
US-09-841-553-5  
; Sequence 5, Application US/09841553  
; Publication No. US20020086402A1  
GENERAL INFORMATION:  
APPLICANT: TAKAKURA, Hikaru  
MORISHITA, Mio  
YAMAMOTO, Katsuhiko  
MIRITA, Masanori  
ASADA, Kiyozo  
TSUNASAWA, Susumu  
KATO, Ikumoshin  
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street N.W., Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/841,553  
FILING DATE: 24-Apr-2001  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/894,818  
FILING DATE: <Unknown>  
APPLICATION NUMBER: JP 323285/1995  
FILING DATE: 12-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TAKAKURA-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 5  
SEQUENCE CHARACTERISTICS:  
LENGTH: 659 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-841-553-5

*Div. of Inpharm Corp.*

Query Match 100.0%; Score 3428; DB 3; Length 659;  
 Best Local Similarity 100.0%; Pred. No. 4,6e-234;  
 Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKGLKALILVILVLGLVGSVAAPPEKKVEQVRNVEKNYGLTPGLFRKIQKLNPNBEIS 60  
 DB 1 MKGLKALILVILVLGLVGSVAAPPEKKVEQVRNVEKNYGLTPGLFRKIQKLNPNBEIS 60  
 QY 61 TVIVFENHREKEIAVAVLELMGAKRVYVHIIPAIADLKVDLIVISGLTGKAKLSGV 120  
 DB 61 TVIVFENHREKEIAVAVLELMGAKRVYVHIIPAIADLKVDLIVISGLTGKAKLSGV 120  
 QY 121 RFIQEDYKTVSALEGLDESAQAQVMAATYVWNLGYDGSGITIGIIDTGIDASHPDLOQKV 180  
 DB 121 RFIQEDYKTVSALEGLDESAQAQVMAATYVWNLGYDGSGITIGIIDTGIDASHPDLOQKV 180  
 QY 181 IGVAVDFVNGRSYFYDDHGHGTHVASIAAGTGAASNGKYGMAPGAKLAGIKVLGADGSGS 240  
 DB 181 IGVAVDFVNGRSYFYDDHGHGTHVASIAAGTGAASNGKYGMAPGAKLAGIKVLGADGSGS 240  
 QY 241 ISTIIGVEMAVVNDKDKYGIKVINLSLGSQSSDGTDSLQAQVNNAMDAGIIVCVAAAGNS 300  
 DB 241 ISTIIGVEMAVVNDKDKYGIKVINLSLGSQSSDGTDSLQAQVNNAMDAGIIVCVAAAGNS 300  
 QY 301 GPTTYTVGSPAAASKYITVGAVDSDNINASFSSRGPTADGRLEPEVVAQVDIIAPRASG 360  
 DB 301 GPTTYTVGSPAAASKYITVGAVDSDNINASFSSRGPTADGRLEPEVVAQVDIIAPRASG 360  
 QY 361 TSMGTPINDYTTKASGSMATPHVSGVGLIIQAHSPMTDPDKYKTLITADIIVAPKEIA 420  
 DB 361 TSMGTPINDYTTKASGSMATPHVSGVGLIIQAHSPMTDPDKYKTLITADIIVAPKEIA 420  
 QY 421 DIAYGAGRNVVYKAIKYDDYAKLFTGSAVDKGSATHTFDVSGATFVTATLYMDTSSDI 480  
 DB 421 DIAYGAGRNVVYKAIKYDDYAKLFTGSAVDKGSATHTFDVSGATFVTATLYMDTSSDI 480  
 QY 481 DLYLDPNGNEVDYSTAYYGFPEKVGYNPTAGTWTVKVSYKGAANYQVDVSDSLSQ 540  
 DB 481 DLYLDPNGNEVDYSTAYYGFPEKVGYNPTAGTWTVKVSYKGAANYQVDVSDSLSQ 540  
 QY 541 SGGGNPNPNPNPPTTDTOTFTGSVNDYDTSDFTMNNSGAKITGDLTFDTSYND 600  
 DB 541 SGGGNPNPNPNPPTTDTOTFTGSVNDYDTSDFTMNNSGAKITGDLTFDTSYND 600  
 QY 601 LDLYLDPNGNLVDRSTSSNSYEHEVYANPAGTWTFLVYASTYGMADYQLKAVVYTG 659  
 DB 601 LDLYLDPNGNLVDRSTSSNSYEHEVYANPAGTWTFLVYASTYGMADYQLKAVVYTG 659

RESULT 2  
 ; Sequence 5, Application US/10800684  
 ; Publication No. US20050014221A1

GENERAL INFORMATION:  
 APPLICANT: TAKAKURA, Hikaru  
 APPLICANT: MORISHITA, Mio  
 APPLICANT: YAMAMOTO, Katsuhiko  
 APPLICANT: MITTA, Masanori  
 APPLICANT: ASADA, Kiyozo  
 APPLICANT: TSUNASAWA, Susumu  
 APPLICANT: KATO, Ikunobu  
 TITLE OF INVENTION: ULTATHERMOSTABLE PROTEASE GENES  
 FILE REFERENCE: TAKAKURA-1A  
 CURRENT APPLICATION NUMBER: US/10/800,684  
 CURRENT FILING DATE: 2004-03-16  
 PRIOR APPLICATION NUMBER: US/09/841,553  
 PRIOR FILING DATE: 2001-04-24  
 PRIOR APPLICATION NUMBER: 08/894,818  
 PRIOR FILING DATE: 1997-08-29  
 PRIOR APPLICATION NUMBER: JP32385/1995  
 PRIOR FILING DATE: 1995-12-12  
 PRIOR APPLICATION NUMBER: JP96/03253

PRIOR FILING DATE: 1996-11-07  
 ; NUMBER OF SEQ ID NOS: 45  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 5

LENGTH: 659  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Synthetic  
 US-10-800-684-5

Query Match 100.0%; Score 3428; DB 5; Length 659;  
 Best Local Similarity 100.0%; Pred. No. 4,6e-234;  
 Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKGLKALILVILVLGLVGSVAAPPEKKVEQVRNVEKNYGLTPGLFRKIQKLNPNBEIS 60  
 DB 1 MKGLKALILVILVLGLVGSVAAPPEKKVEQVRNVEKNYGLTPGLFRKIQKLNPNBEIS 60  
 QY 61 TVIVFENHREKEIAVAVLELMGAKRVYVHIIPAIADLKVDLIVISGLTGKAKLSGV 120  
 DB 61 TVIVFENHREKEIAVAVLELMGAKRVYVHIIPAIADLKVDLIVISGLTGKAKLSGV 120  
 QY 121 RFIQEDYKTVSALEGLDESAQAQVMAATYVWNLGYDGSGITIGIIDTGIDASHPDLOQKV 180  
 DB 121 RFIQEDYKTVSALEGLDESAQAQVMAATYVWNLGYDGSGITIGIIDTGIDASHPDLOQKV 180  
 QY 181 IGVAVDFVNGRSYFYDDHGHGTHVASIAAGTGAASNGKYGMAPGAKLAGIKVLGADGSGS 240  
 DB 181 IGVAVDFVNGRSYFYDDHGHGTHVASIAAGTGAASNGKYGMAPGAKLAGIKVLGADGSGS 240  
 QY 241 ISTIIGVEMAVVNDKDKYGIKVINLSLGSQSSDGTDSLQAQVNNAMDAGIIVCVAAAGNS 300  
 DB 241 ISTIIGVEMAVVNDKDKYGIKVINLSLGSQSSDGTDSLQAQVNNAMDAGIIVCVAAAGNS 300  
 QY 301 GPTTYTVGSPAAASKYITVGAVDSDNINASFSSRGPTADGRLEPEVVAQVDIIAPRASG 360  
 DB 301 GPTTYTVGSPAAASKYITVGAVDSDNINASFSSRGPTADGRLEPEVVAQVDIIAPRASG 360  
 QY 421 DIAYGAGRNVVYKAIKYDDYAKLFTGSAVDKGSATHTFDVSGATFVTATLYMDTSSDI 480  
 DB 421 DIAYGAGRNVVYKAIKYDDYAKLFTGSAVDKGSATHTFDVSGATFVTATLYMDTSSDI 480  
 QY 481 DLYLDPNGNEVDYSTAYYGFPEKVGYNPTAGTWTVKVSYKGAANYQVDVSDSLSQ 540  
 DB 481 DLYLDPNGNEVDYSTAYYGFPEKVGYNPTAGTWTVKVSYKGAANYQVDVSDSLSQ 540  
 QY 541 SGGGNPNPNPNPPTTDTOTFTGSVNDYDTSDFTMNNSGAKITGDLTFDTSYND 600  
 DB 541 SGGGNPNPNPNPPTTDTOTFTGSVNDYDTSDFTMNNSGAKITGDLTFDTSYND 600  
 QY 601 LDLYLDPNGNLVDRSTSSNSYEHEVYANPAGTWTFLVYASTYGMADYQLKAVVYTG 659  
 DB 601 LDLYLDPNGNLVDRSTSSNSYEHEVYANPAGTWTFLVYASTYGMADYQLKAVVYTG 659

RESULT 3  
 ; Sequence 35, Application US/09841553  
 ; Publication No. US20020086402A1

GENERAL INFORMATION:  
 APPLICANT: TAKAKURA, Hikaru  
 APPLICANT: MORISHITA, Mio  
 APPLICANT: YAMAMOTO, Katsuhiko  
 APPLICANT: MITTA, Masanori  
 APPLICANT: ASADA, Kiyozo  
 APPLICANT: TSUNASAWA, Susumu  
 APPLICANT: KATO, Ikunobu  
 TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES



```

NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSER: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/841,553
FILING DATE: 24-Apr-2001

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/894,818
FILING DATE: <Unknown>
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995

ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528

INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 654 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-09-841-553-35

Query Match      87.4%; Score 2997.5; DB 3; Length 654;
Best Local Similarity 88.3%; Pred. No. 1.5e-203;
Matches 582; Conservative 22; Mismatches 50; Indels 5; Gaps 1;

QY 1 MGKLKLLIIVILGLVGSVAAAPKKYEQVNRVEKNYGLTPGLFRKIQKLNPEEIS 60
DB 1 MGKLKLLIIVILGLVGSVAAAPKKYEQVNRVEKNYGLTPGLFRKIQKLNPEEIS 60
QY 61 TVIVFENHREKEIAVRLVLEMGAKRYVYHIIIPALADLKVRDLVLSGLTGSKALSGV 120
DB 61 TVIVFENHREKEIAVRLVLEMGAKRYVYHIIIPALADLKVRDLVLSGLTGSKALSGV 120
QY 121 RFIQEDYKYTVSALEGLDESAQAQWATYVMNLGYDGSGITIGIIDTGIDASHPDLOGKY 180
DB 121 RFIQEDYKYTVSALEGLDESAQAQWATYVMNLGYDGSGITIGIIDTGIDASHPDLOGKY 180
QY 181 IGVWDFVNRASYPYDDHGHGTHVASTAAGTGAASNGKYGMAPGALAIKYLGAAGSSG 240
DB 181 IGVWDFVNRASYPYDDHGHGTHVASTAAGTGAASNGKYGMAPGALAIKYLGAAGSSG 240
QY 181 IGVWDFVNRASYPYDDHGHGTHVASTAAGTGAASNGKYGMAPGALAIKYLGAAGSSG 240
DB 181 IGVWDFVNRASYPYDDHGHGTHVASTAAGTGAASNGKYGMAPGALAIKYLGAAGSSG 240
QY 241 ISTIIKGVMAVDNKKYGIKYNLSLSSQSSDGTDSLQAVNNAMDAIGIVVCAAGNS 300
DB 241 ISTIIKGVMAVDNKKYGIKYNLSLSSQSSDGTDSLQAVNNAMDAIGIVVCAAGNS 300
QY 241 ISTIIKGVMAVDNKKYGIKYNLSLSSQSSDGTDSLQAVNNAMDAIGIVVCAAGNS 300
DB 241 ISTIIKGVMAVDNKKYGIKYNLSLSSQSSDGTDSLQAVNNAMDAIGIVVCAAGNS 300
QY 301 GPNYTVVGSPPAAASKYITVGAVDSDNINASFSSRGPTADGRLEPEVAVPAGNMIIAPRAG 360
DB 301 GPNYTVVGSPPAAASKYITVGAVDSDNINASFSSRGPTADGRLEPEVAVPAGNMIIAPRAG 360
QY 361 TSMGPIINDYYTKASGTSMTPEHVSQVGLIIQAHPSWTPDKYKTLIETADIAPKEIA 420
DB 361 TSMGPIINDYYTKASGTSMTPEHVSQVGLIIQAHPSWTPDKYKTLIETADIAPKEIA 420
QY 421 DIAYGAGRNVVYKAIKYDDYAKLTFTGVSADKGSATHTFVSGATFVTATLYWDTSSDI 480
DB 421 DIAYGAGRNVVYKAIKYDDYAKLTFTGVSADKGSATHTFVSGATFVTATLYWDTSSDI 480
```

```

QY 481 DLYLXDPNGNEVDYSTAYTAGEKGYNPNATGATWTKVYSKGAANYQVDVSDSLQ 540
DB 481 DLYLXDPNGNEVDYSTAYTAGEKGYNPNATGATWTKVYSKGAANYQVDVSDSLQ 540
QY 541 SGGGNPNPNPNPPTTDTQFTGSDVNDYMDTSDTFMNVNSGATKIGDLTFDTSYND 600
DB 541 SGGGNPNPNPNPPTTDTQFTGSDVNDYMDTSDTFMNVNSGATKIGDLTFDTSYND 600
QY 541 SGGGNPNPNPNPPTTDTQFTGSDVNDYMDTSDTFMNVNSGATKIGDLTFDTSYND 600
DB 541 SGGGNPNPNPNPPTTDTQFTGSDVNDYMDTSDTFMNVNSGATKIGDLTFDTSYND 600
QY 601 LDLYLXDPNGNLYVDRSTSSNYSHEVYANAPAGTWTFLYAYASTGYMADYOLKAVVY 659
DB 601 LDLYLXDPNGNLYVDRSTSSNYSHEVYANAPAGTWTFLYAYASTGYMADYOLKAVVY 659
QY 659 LDLYLXDPNGNLYVDRSTSSNYSHEVYANAPAGTWTFLYAYASTGYMADYOLKAVVY 659
DB 659 LDLYLXDPNGNLYVDRSTSSNYSHEVYANAPAGTWTFLYAYASTGYMADYOLKAVVY 659

RESULT 4
US-10-090-624-16
Sequence 16, Application US/10090624
Publication No. US20020132335A1
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hi-karu
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Kunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA=6
CURRENT APPLICATION NUMBER: US/10/090,624
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: 09/445,472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 16
LENGTH: 654
TYPE: PRT
ORGANISM: Pyrococcus furiosus
US-10-090-624-16

Query Match      87.4%; Score 2997.5; DB 4; Length 654;
Best Local Similarity 88.3%; Pred. No. 1.5e-203;
Matches 582; Conservative 22; Mismatches 50; Indels 5; Gaps 1;

QY 1 MGKLKLLIIVILGLVGSVAAAPKKYEQVNRVEKNYGLTPGLFRKIQKLNPEEIS 60
DB 1 MGKLKLLIIVILGLVGSVAAAPKKYEQVNRVEKNYGLTPGLFRKIQKLNPEEIS 60
QY 61 TVIVFENHREKEIAVRLVLEMGAKRYVYHIIIPALADLKVRDLVLSGLTGSKALSGV 120
DB 61 TVIVFENHREKEIAVRLVLEMGAKRYVYHIIIPALADLKVRDLVLSGLTGSKALSGV 120
QY 121 RFIQEDYKYTVSALEGLDESAQAQWATYVMNLGYDGSGITIGIIDTGIDASHPDLOGKY 180
DB 121 RFIQEDYKYTVSALEGLDESAQAQWATYVMNLGYDGSGITIGIIDTGIDASHPDLOGKY 180
QY 181 IGVWDFVNRASYPYDDHGHGTHVASTAAGTGAASNGKYGMAPGALAIKYLGAAGSSG 240
DB 181 IGVWDFVNRASYPYDDHGHGTHVASTAAGTGAASNGKYGMAPGALAIKYLGAAGSSG 240
QY 181 IGVWDFVNRASYPYDDHGHGTHVASTAAGTGAASNGKYGMAPGALAIKYLGAAGSSG 240
DB 181 IGVWDFVNRASYPYDDHGHGTHVASTAAGTGAASNGKYGMAPGALAIKYLGAAGSSG 240
QY 241 ISTIIKGVMAVDNKKYGIKYNLSLSSQSSDGTDSLQAVNNAMDAIGIVVCAAGNS 300
DB 241 ISTIIKGVMAVDNKKYGIKYNLSLSSQSSDGTDSLQAVNNAMDAIGIVVCAAGNS 300
QY 241 ISTIIKGVMAVDNKKYGIKYNLSLSSQSSDGTDSLQAVNNAMDAIGIVVCAAGNS 300
DB 241 ISTIIKGVMAVDNKKYGIKYNLSLSSQSSDGTDSLQAVNNAMDAIGIVVCAAGNS 300
QY 301 GPNYTVVGSPPAAASKYITVGAVDSDNINASFSSRGPTADGRLEPEVAVPAGNMIIAPRAG 360
DB 301 GPNYTVVGSPPAAASKYITVGAVDSDNINASFSSRGPTADGRLEPEVAVPAGNMIIAPRAG 360
QY 361 TSMGPIINDYYTKASGTSMTPEHVSQVGLIIQAHPSWTPDKYKTLIETADIAPKEIA 420
DB 361 TSMGPIINDYYTKASGTSMTPEHVSQVGLIIQAHPSWTPDKYKTLIETADIAPKEIA 420
QY 421 DIAYGAGRNVVYKAIKYDDYAKLTFTGVSADKGSATHTFVSGATFVTATLYWDTSSDI 480
DB 421 DIAYGAGRNVVYKAIKYDDYAKLTFTGVSADKGSATHTFVSGATFVTATLYWDTSSDI 480
```

```

Db 421 DIAAGARVAAKAIANDYNAKLVFTGVANKSGQTHQFVLSGASFTATILYMDNANSDL 480
Qy 481 DLYLYDNGNEVDYSYATAYGFEKVGYNPTAGTWITKVYSYKAAANYQVDVSDGSLQ 540
Db 481 DLYLYDNGNOVDYSYATAYGFEKVGYNPTDGTWITKVYSYSGSANYQVDVSDGSLQ 540
Qy 541 SGGGNPNPNPNPTPTDTCFTGVSVDYWDTSDFTFMNVNSGATKITGDLTFPTSYND 600
Db 541 -----PGSSPSPQPEPTVDAKTFQGSDDHYVDRSDFTMTVNSGATKITGDLTFPTSYND 595
Qy 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTLVYASTYGMADYQLKAVVYVG 659
Db 596 LDLYLYDPNGNLVDRSTSSNSYEHVEYLTAPGTWTLVYASTYGMADYQLKAVVYVG 654

```

# RESULT 5 US-10-800-684-35

```

/ Sequence 35, Application US/10800684
/ Publication No. US20050014221A1
/ GENERAL INFORMATION:
/ APPLICANT: TAKAKURA, Hikaru
/ APPLICANT: MORISHITA, Mio
/ APPLICANT: YAMAMOTO, Katsuhiko
/ APPLICANT: MITTA, Masaoori
/ APPLICANT: ASADA, Kiyozo
/ APPLICANT: TSUNASAWA, Susumu
/ APPLICANT: KATO, Ikunoshin
/ TITLE OF INVENTION: ULTRATHERMOSTABLE PROTEASE GENES
/ FILE REFERENCE: TAKAKURA-1A
/ CURRENT APPLICATION NUMBER: US/10/800,684
/ PRIOR FILING DATE: 2004-03-16
/ PRIOR APPLICATION NUMBER: US/09/841,553
/ PRIOR FILING DATE: 2001-04-24
/ PRIOR APPLICATION NUMBER: 08/894,818
/ PRIOR FILING DATE: 1997-08-29
/ PRIOR APPLICATION NUMBER: JP3285/1995
/ PRIOR FILING DATE: 1995-12-12
/ PRIOR APPLICATION NUMBER: JP96/03253
/ PRIOR FILING DATE: 1996-11-07
/ NUMBER OF SEQ ID NOS: 45
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 35
/ LENGTH: 654
/ TYPE: PRT
/ ORGANISM: Pyrococcus furiosus
/ US-10-800-684-35

```

Query Match 87.4%; Score 2997.5; DB 5; Length 654;  
Best Local Similarity 88.3%; Pred. No. 1.5e-203;  
Matches 582; Conservative 22; Mismatches 50; Indels 5; Gaps 1;

```

Qy 1 MGKGLKLLIIVLVLGLVVGSAVAAPEKKYBOVRNVEKNYGLTTPGLFRKIQKLNPNERS 60
Db 1 MGKGLKLLIIVLVLGLVVGSAVAAPEKKYBOVRNVEKNYGLTTPGLFRKIQKLNPNERS 60
Qy 61 TYIVFENHREKEIAVRVLELGMKAVRYVYHIIIPAIADLKVRDLVLSGLTGKAKLSGV 120
Db 61 TYIVFENHREKEIAVRVLELGMKAVRYVYHIIIPAIADLKVRDLVLSGLTGKAKLSGV 120
Qy 121 RFIQEDYKVTYSABELGIDESAQVMAATYVNNLGYDSGITTIGIIDGIDASHPDLQKV 180
Db 121 RFIQEDYKVTYSABELGIDESAQVMAATYVNNLGYDSGITTIGIIDGIDASHPDLQKV 180
Qy 181 IGMVDFVNGRSYYPDDHGHGTHVASIAAGTGAASNGKYGMAPGAAGIIVVCAAGNS 240
Db 181 IGMVDFVNGRSYYPDDHGHGTHVASIAAGTGAASNGKYGMAPGAAGIIVVCAAGNS 240
Qy 241 ISTIIGVEMAVDNKDKYGIKVINISLSSQSSDGTDSLQAQVNNAMDAIGIVVCAAGNS 300
Db 241 ISTIIGVEMAVDNKDKYGIKVINISLSSQSSDGTDSLQAQVNNAMDAIGIVVCAAGNS 300
Qy 301 GNTTYVGSPPAAASKYITTGAVDSNDNITASFSSRGFTADGRLKPEVVAAGVDITIAPRASG 360

```

```

Db 301 GPKKTYITGSPAAASKYITTGAVDKYVITSFSSRGFTADGRLKPEVVAAGNMIIAPRASG 360
Qy 361 TSMGTPDINDYTTAASGTSMATPHVSGVALLIQAPBSWTPDKYKTALETADIAPKETA 420
Db 361 TSMGQPIINDYTTAAPTSMATPHVAGIAALLIQAPBSWTPDKYKTALETADIAPKETA 420
Qy 421 DIAAGARVAAKAIANDYNAKLVFTGVANKSGQTHQFVLSGASFTATILYMDNANSDL 480
Db 421 DIAAGARVAAKAIANDYNAKLVFTGVANKSGQTHQFVLSGASFTATILYMDNANSDL 480
Qy 481 DLYLYDNGNEVDYSYATAYGFEKVGYNPTAGTWITKVYSYKAAANYQVDVSDGSLQ 540
Db 481 DLYLYDNGNOVDYSYATAYGFEKVGYNPTDGTWITKVYSYSGSANYQVDVSDGSLQ 540
Qy 541 SGGGNPNPNPNPTPTDTCFTGVSVDYWDTSDFTFMNVNSGATKITGDLTFPTSYND 600
Db 541 -----PGSSPSPQPEPTVDAKTFQGSDDHYVDRSDFTMTVNSGATKITGDLTFPTSYND 595
Qy 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTLVYASTYGMADYQLKAVVYVG 659
Db 596 LDLYLYDPNGNLVDRSTSSNSYEHVEYLTAPGTWTLVYASTYGMADYQLKAVVYVG 654

```

# RESULT 6 US-10-888-588-16

```

/ Sequence 16, Application US/10888588
/ Publication No. US20050084934A1
/ GENERAL INFORMATION:
/ APPLICANT: TAKAKURA, Hikaru
/ APPLICANT: MORISHITA, Mio
/ APPLICANT: SHIMOJO, Tomoko
/ APPLICANT: ASADA, Kiyozo
/ APPLICANT: KATO, Ikunoshin
/ TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
/ FILE REFERENCE: TAKAKURA-6
/ CURRENT APPLICATION NUMBER: US/10/888,588
/ PRIOR FILING DATE: 2004-07-12
/ PRIOR APPLICATION NUMBER: US/10/090,624
/ PRIOR FILING DATE: 2002-03-06
/ PRIOR APPLICATION NUMBER: 09/445,472
/ PRIOR FILING DATE: 1999-12-06
/ PRIOR APPLICATION NUMBER: 151969/1997
/ PRIOR FILING DATE: 1997-06-10
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 16
/ LENGTH: 654
/ TYPE: PRT
/ ORGANISM: Pyrococcus furiosus
/ US-10-888-588-16

```

Query Match 87.4%; Score 2997.5; DB 5; Length 654;  
Best Local Similarity 88.3%; Pred. No. 1.5e-203;  
Matches 582; Conservative 22; Mismatches 50; Indels 5; Gaps 1;

```

Qy 1 MGKGLKLLIIVLVLGLVVGSAVAAPEKKYBOVRNVEKNYGLTTPGLFRKIQKLNPNERS 60
Db 1 MGKGLKLLIIVLVLGLVVGSAVAAPEKKYBOVRNVEKNYGLTTPGLFRKIQKLNPNERS 60
Qy 61 TYIVFENHREKEIAVRVLELGMKAVRYVYHIIIPAIADLKVRDLVLSGLTGKAKLSGV 120
Db 61 TYIVFENHREKEIAVRVLELGMKAVRYVYHIIIPAIADLKVRDLVLSGLTGKAKLSGV 120
Qy 121 RFIQEDYKVTYSABELGIDESAQVMAATYVNNLGYDSGITTIGIIDGIDASHPDLQKV 180
Db 121 RFIQEDYKVTYSABELGIDESAQVMAATYVNNLGYDSGITTIGIIDGIDASHPDLQKV 180
Qy 181 IGMVDFVNGRSYYPDDHGHGTHVASIAAGTGAASNGKYGMAPGAAGIIVVCAAGNS 240
Db 181 IGMVDFVNGRSYYPDDHGHGTHVASIAAGTGAASNGKYGMAPGAAGIIVVCAAGNS 240
Qy 241 ISTIIGVEMAVDNKDKYGIKVINISLSSQSSDGTDSLQAQVNNAMDAIGIVVCAAGNS 300

```

```

Db      241  ISIIIGVEAWADNKKQYGIKVINISLSGSSGSDTDLSQAVNAAMADGLVYVVAAGNS 3000
Qy      301  GPNITTVGSPAAASKVITVGAVDSDNNTASFSSRGPTADGRKEVYAPGYDIIAPRAG 3606
Db      301  GPKHYITIGSPAAASKVITVGAVDKVDVITSFSSRGPTADGRKEVYAPGWMIIAADASG 3606
Qy      361  TSMGTEPINDYYTKASTSMATPHVSGVCAIIIOAHPSWTPDKVKTALILETDIYAPKEIA 4220
Db      361  TSMGQEPINDYYTAAFTSMATPHVAGIAALLIOAHPSWTPDKVKTALILETDIYKPEEIA 4200
Qy      421  DIAYGGRWVYKAIKCYDDYAKLTPTGSAVDGSAHTHFVDSGATFVYATLTVMPTGSSDI 4800
Db      421  DIAYGGRWYAKAIYNDYAKLTFTGVYANGSQTHQFVJSGASFVYATLTVMYMANASDL 4800
Qy      481  DLYLYDPNGNEVDYSTAYYGEKGYVNPATGTTVYVSKGANAQVVDVSDGSLSQ 5400
Db      481  DLYLYDPNGNAQVYSTAYYGEKGYVNPDTGVTIKVYSISGSANQVVDVSDGSLSQ 5400
Qy      541  SGGAGNPENPNPPTPTDQTQTFQSGVNDYMDTSDPTFMANNVSGATKTIGDLTFDTSYND 6000
Db      541  -----PGSSPSPQPEPTVDAKTIFQSGDHYIYDRSPTFMVYNSGATKTIGDLVVDTSYHD 5950
Qy      601  LDLYLYDPNGNLVDNRSTSSNSYEHEVYANPAEGTFTFLVYAVSTYGMADYOLKAVVYVG 6559
Db      596  LDLYLYDPNQKLVDRSENSYEHVYELTPAGCTYFLVYAVYITGYMAVYELTKAVYVG 654

```

RESULT 7  
 US-09-841-553-1  
 Sequence 1, Application US/09841553  
 Publication No. US20020086402A1  
 GENERAL INFORMATION:  
 APPLICANT: TAKAKURA, HiKaru  
 MORISHITA, Mio  
 YAMAMOTO, Katsuhiko  
 MITTA, Masanori  
 ASADA, Kiyozo  
 TSUNASAWA, Susumu  
 KATO, Ikumoshin  
 TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
 NUMBER OF SEQUENCES: 42  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Browdy and Neimark  
 STREET: 419 Seventh Street N.W., Ste. 300  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: United States of America  
 ZIP: 20004  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/841,553  
 FILING DATE: 24-Apr-2001  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/894,818  
 FILING DATE: <Unknown>  
 APPLICATION NUMBER: JP 323285/1995  
 FILING DATE: 12-DEC-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Browdy, Roger L.  
 REGISTRATION NUMBER: 25, 618  
 REFERENCE/DOCKET NUMBER: TAKAKURA=1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 628-5197  
 TELEFAX: (202) 737-3528.  
 INFORMATION FOR SEQ ID NO. 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 659 amino acids  
 TYPE: amino acid

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-841-553-1

Query Match      85.0%; Score 2914; DB 3; Length 659;
Query Local Similarity 83.7%; Pred. No. 1,2e-197;
Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps 7.

```

|    |     |  |     |
|----|-----|--|-----|
| Qy | 1   | MKGJKALILVILVGLVWGSVAAPAKKKVEBOVRN---VEKNYGGILTPELFKRIQOLQNE   | 57  |
| Db | 1   | MKRLGAVLVLVLVGLAGLTALAPKVPV--VRNNAVOQKNYGIITLPELFKVOGRMAMO     | 58  |
| Qy | 58  | EISTFVIFENREKEIAVRLEIMGAKRVYTHIIPALADLKVRDLVIGSL--TG--G        | 113 |
| Db | 59  | EVDFTVMEGSGYDRBRAVKVRLMLGAQCKVYSYKIIIPAAVAKIKARDLLILAGMIDTGYFG | 118 |
| Qy | 114 | KAKLSGRFTIOEDKYKTVYSALBEGLDSEAQVMAITYWN--LGYGSGTITIGIDGTIDAS   | 172 |
| Db | 119 | NTRVSGIKFIQEDKYKVOVDDA----TSVSGIADTYWNSLIGYDGSQVVAIVDTGIDAN    | 173 |
| Qy | 173 | HPDLQGVKIVGVDVPVNGRSYPYDDHGHGTHVASIAGTGAASNGKRYKMAAGAKIAGIKV   | 232 |
| Db | 174 | HPDLKGVKIGVYDAVNGRSTPYDQGHGTHVAGIYVAGTSV--NSQYIGVAPGAKIKVGVK   | 232 |
| Qy | 233 | LGAAGSGSISTIIKGVEMAVDNKRGYGIKVINLSIGSSQSSDGTDSLSEAANNAMDAIGV   | 292 |
| Db | 233 | LGAAGSGSVSTIIAGVDMVVOVKXKRGIRVINLSIGSSQSSDGTDSLSEAANNAMDAIGV   | 292 |
| Qy | 293 | VCVAAGNSGENTYVGSPPAAASKVITYGAVDSDNDNIASFSSRGPTADGRLKPEVVAQVD   | 352 |
| Db | 293 | VCVAAGNSGPTTYVGSPPAAASKIITVGAVDSDNNIASFSSRGPTADGRLKPEVVAQVD    | 352 |
| Qy | 353 | IIAPRAGTSMGTINDYYTKASTSMATPHVSGVALLIQAHSWTPDKKXTALILETAD       | 412 |
| Db | 353 | IIAPRAGTSMGTINDYYTKASGSMATPHVSGVALLIQAHSWTPDKKXTALILETAD       | 412 |
| Qy | 413 | IVAEREIADIIYAGRVVVKAIKYDDPAKLTFTGSVADKSAHTFDFVSGATFVATILY      | 472 |
| Db | 413 | IVAEREIADIIYAGRVVVKAIKYDDPAKLTFTGSVADKSAHTTTFVSGATFVATILY      | 472 |
| Qy | 473 | WDTGSSSIDLYLDPNGNEVDYSYTAAYGFEKGVYNFTAGTWTVKVSYKGAANYQVDV      | 532 |
| Db | 473 | WDTGSSSIDLYLDPNGNEVDYSYTAAYGFEKGVYNFTAGTWTVKVSYKGAANYQVDV      | 532 |
| Qy | 533 | VSDGSLSGSGGAGNPENPNPPTTDTQTFTGSVNDYWDTSDFITMNVNSGATKITGDL      | 592 |
| Db | 533 | VSDGSLSGSGGAGNPENPNPPTTDTQTFTGSVNDYWDTSDFITMNVNSGATKITGDL      | 592 |
| Qy | 593 | TFDTSVNDLIDLXLDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFIYYASTYGMADYOL    | 652 |
| Db | 593 | TFDTSVNDLIDLXLDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFIYYASTYGMADYOL    | 652 |
| Qy | 653 | KAVVYYG 659  |     |
| Db | 653 | KAVVYYG 659  |     |

RESULT 8  
US-10-090-624-12  
: Sequence 12, Application US/10090624  
: Publication No. US2002013335A1  
: GENERAL INFORMATION:  
: APPLICANT: TAKAKURA, HiKaru  
: APPLICANT: MORISHITA, Mio  
: APPLICANT: SHIMOTO, Tomoko  
: APPLICANT: ASADA, Kiyozo  
: APPLICANT: KATO, Ikumoshin  
: TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
: FILE REFERENCE: TAKAKURA=6  
: CURRENT APPLICATION NUMBER: US/10/090,624  
: CURRENT FILING DATE: 2002-03-06

; PRIOR APPLICATION NUMBER: 09/445,472  
 ; PRIOR FILING DATE: 1999-12-06  
 ; PRIOR APPLICATION NUMBER: 151969/1997  
 ; PRIOR FILING DATE: 1997-06-10  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 12  
 ; LENGTH: 659  
 ; TYPE: PR  
 ; ORGANISM: Thermococcus celer  
 US-10-090-624-12

Query Match 85.0%; Score 2914; DB 4; Length 659;  
 Best Local Similarity 83.7%; Pred. No. 1.2e-197;  
 Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps 7;

```

QY 1 MKGLKALILVILVGLVGVSAAPAEKKEVQVRN---VEKNYGLLTGGLFRKIQKLNPE 57
DB 1 MKRLGAVVLAIVLVLGGLAGTALAAPVPRV--VRNNAVOQKNYGLLTGGLFRKVOQRMNMQ 58
QY 58 EISTVIVFENHREKEIAVRVLELMGAKVRYVYHIIIPAIADLKVRDLVITSGL--TG--G 113
DB 59 EVDIVIMFGSYGRDRRAVKVLRMGAGVYKSYKIIIPAAVKIKARLDLLIAGMIDTGYFG 118
QY 114 KAKISGVRFIOEDYKVTVAELBGLDESAQVMATVYWN--LGYDGSGITIGIITDGDAS 172
DB 119 NTRVSGIKRFOEDYKQVDDA-----TSVQIGADVTWNSLGYDGSVVVAIVDTGIDAN 173
QY 173 HPDLGKVIQWVDPVNGRSYPYDDHGHGTHVASIAGTGAASNGKYKGMAPGAKLAGIKY 232
DB 174 HPDLGKVIQWDAVNGRSTPYDDHGHGTHVAGIAGTGSV--NSQYIGVAPGAKLVGVKY 232
QY 233 LGADSGSISTIIKGVEMAVNDKDKYGIKVINLSLSSQSSDGTDSLQAANNMADGIV 292
DB 233 LGADSGSVSTIIAGVDWVQNKDKYGIKVINLSLSSQSSDGTDSLQAANNMADGIV 292
QY 293 VCVAAAGSGPNTYTVGSPAAASKVITVGAVDSDNNTIASFSSRGPTADGRLKPEVAPGV 352
DB 293 VCVAAAGSGPNTYTVGSPAAASKVITVGAVDSDNNTIASFSSRGPTADGRLKPEVAPGV 352
QY 353 IIPRAGSGTSGTPIINDYTTKASGTSMAATPHVSGVGLIIQHPBMTDPDKXTALIEFAD 412
DB 353 IIPRAGSGTSGTPIINDYTTKASGTSMAATPHVSGVGLIIQHPBMTDPDKXTALIEFAD 412
QY 413 IYAPKEIADIAYAGRVNVYKAIKYDDYAKLFTGSAVDKGSATHTFEDVSGATFVATLY 472
DB 413 IYAPKEIADIAYAGRVNVYKAIKYDDYAKLFTGSAVDKGSATHTFEDVSGATFVATLY 472
QY 473 WDTGSSDIDLILYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTIVKVSYKGAANYQVDV 532
DB 473 WDTGSSDIDLILYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTIVKVSYKGAANYQVDV 532
QY 533 VSDGSLSGSGGPNPNPNPTPTDQFTGTVNDYDTSPTFMNPNVNSGATKKTIGDL 592
DB 533 VSDGSLSGSGGPNPNPNPTPTDQFTGTVNDYDTSPTFMNPNVNSGATKKTIGDL 592
QY 593 TFDTSYNDLIDLILYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVLYASTYGMADYQL 652
DB 593 TFDTSYNDLIDLILYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVLYASTYGMADYQL 652
QY 653 KAVVYVG 659
DB 653 KAVVYVG 659
  
```

RESULT 9  
 US-10-800-684-1

; Sequence 1, Application US/10800684  
 ; Publication No. US20050014221A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TAKAKURA, Hikaru  
 ; APPLICANT: MORISHITA, Mio  
 ; APPLICANT: YAMAMOTO, Katsumiko

; APPLICANT: MITTA, Masanori  
 ; APPLICANT: ASADA, Kiyozo  
 ; APPLICANT: TSUNASAWA, Susumu  
 ; APPLICANT: KATO, Ikumoshin  
 ; TITLE OF INVENTION: ULTRATHERMOSTABLE PROTEASE GENES  
 ; FILE REFERENCE: TAKAKURA=1A  
 ; CURRENT APPLICATION NUMBER: US/10/800,684  
 ; PRIOR FILING DATE: 2004-03-16  
 ; PRIOR APPLICATION NUMBER: US/09/841,553  
 ; PRIOR FILING DATE: 2001-04-24  
 ; PRIOR APPLICATION NUMBER: 08/894,818  
 ; PRIOR FILING DATE: 1997-08-29  
 ; PRIOR APPLICATION NUMBER: JP32385/1995  
 ; PRIOR FILING DATE: 1995-12-12  
 ; PRIOR APPLICATION NUMBER: JP96/03253  
 ; PRIOR FILING DATE: 1996-11-07  
 ; NUMBER OF SEQ ID NOS: 45  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 1  
 ; LENGTH: 659  
 ; TYPE: PR  
 ; ORGANISM: Thermococcus celer  
 US-10-800-684-1

Query Match 85.0%; Score 2914; DB 5; Length 659;  
 Best Local Similarity 83.7%; Pred. No. 1.2e-197;  
 Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps 7;

```

QY 1 MKGLKALILVILVGLVGVSAAPAEKKEVQVRN---VEKNYGLLTGGLFRKIQKLNPE 57
DB 1 MKRLGAVVLAIVLVLGGLAGTALAAPVPRV--VRNNAVOQKNYGLLTGGLFRKVOQRMNMQ 58
QY 58 EISTVIVFENHREKEIAVRVLELMGAKVRYVYHIIIPAIADLKVRDLVITSGL--TG--G 113
DB 59 EVDIVIMFGSYGRDRRAVKVLRMGAGVYKSYKIIIPAAVKIKARLDLLIAGMIDTGYFG 118
QY 114 KAKISGVRFIOEDYKVTVAELBGLDESAQVMATVYWN--LGYDGSGITIGIITDGDAS 172
DB 119 NTRVSGIKRFOEDYKQVDDA-----TSVQIGADVTWNSLGYDGSVVVAIVDTGIDAN 173
QY 173 HPDLGKVIQWVDPVNGRSYPYDDHGHGTHVASIAGTGAASNGKYKGMAPGAKLAGIKY 232
DB 174 HPDLGKVIQWDAVNGRSTPYDDHGHGTHVAGIAGTGSV--NSQYIGVAPGAKLVGVKY 232
QY 233 LGADSGSISTIIKGVEMAVNDKDKYGIKVINLSLSSQSSDGTDSLQAANNMADGIV 292
DB 233 LGADSGSVSTIIAGVDWVQNKDKYGIKVINLSLSSQSSDGTDSLQAANNMADGIV 292
QY 293 VCVAAAGSGPNTYTVGSPAAASKVITVGAVDSDNNTIASFSSRGPTADGRLKPEVAPGV 352
DB 293 VCVAAAGSGPNTYTVGSPAAASKVITVGAVDSDNNTIASFSSRGPTADGRLKPEVAPGV 352
QY 353 IIPRAGSGTSGTPIINDYTTKASGTSMAATPHVSGVGLIIQHPBMTDPDKXTALIEFAD 412
DB 353 IIPRAGSGTSGTPIINDYTTKASGTSMAATPHVSGVGLIIQHPBMTDPDKXTALIEFAD 412
QY 413 IYAPKEIADIAYAGRVNVYKAIKYDDYAKLFTGSAVDKGSATHTFEDVSGATFVATLY 472
DB 413 IYAPKEIADIAYAGRVNVYKAIKYDDYAKLFTGSAVDKGSATHTFEDVSGATFVATLY 472
QY 473 WDTGSSDIDLILYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTIVKVSYKGAANYQVDV 532
DB 473 WDTGSSDIDLILYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTIVKVSYKGAANYQVDV 532
QY 533 VSDGSLSGSGGPNPNPNPTPTDQFTGTVNDYDTSPTFMNPNVNSGATKKTIGDL 592
DB 533 VSDGSLSGSGGPNPNPNPTPTDQFTGTVNDYDTSPTFMNPNVNSGATKKTIGDL 592
QY 593 TFDTSYNDLIDLILYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVLYASTYGMADYQL 652
DB 593 TFDTSYNDLIDLILYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVLYASTYGMADYQL 652
QY 653 KAVVYVG 659
DB 653 KAVVYVG 659
  
```

Db 653 KAVVYVG 659

RESULT 10

US-10-888-588-12

Sequence 12, Application US/10888588

Publication No. US20050084934A1

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru

APPLICANT: MORISHITA, Mio

APPLICANT: SHIMOTO, Tomoko

APPLICANT: ASADA, Kiyozo

APPLICANT: KATO, Ikunoshin

TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE

FILE REFERENCE: TAKAKURA-6

CURRENT APPLICATION NUMBER: US/10/888,588

CURRENT FILING DATE: 2004-07-12

PRIOR APPLICATION NUMBER: US/10/090,624

PRIOR FILING DATE: 2002-03-06

PRIOR APPLICATION NUMBER: 09/445,472

PRIOR FILING DATE: 1999-12-06

PRIOR APPLICATION NUMBER: 151969/1997

PRIOR FILING DATE: 1997-06-10

NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn version 3.0

SEQ ID NO 12

LENGTH: 659

TYPE: PRT

ORGANISM: Thermococcus celer

US-10-888-588-12

Query Match

Best Local Similarity 85.0%; Score 2914; DB 5; Length 659;

Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps 7;

1 MKGLKALIVIVLGLVGSVAAPAKKEQVYRN--VEKNYGLTLPGLFRKIQKLNPN 57

1 MRLGLAVNLAIVGLAGTALAAPYKPV--VRNNAVQKNGTGLTPGLFKYQVRNNMG 58

58 EISTVIFENHREKIAVLEIMGAKRVYVYHIIIPALADLKVRDLVLSGL--TG--G 113

59 EVDTVIMFGSYGRDRAVKVRLMGAQVYKYSKIPAVAVKIKARDLLIAGMIDGYRG 118

114 KAKLSVRIQEDRYKTVSAELEGLDESAQVMATYVNM--LGYDGSGITIGIIDGIDS 172

119 NTRVSGIKRFOEDYKQVDA-----TSVQIGADVWNSLGYDGSVVVAIVDTGIDAN 173

173 HPDLQKVIQWVDFVNGRSPYDDHGHVHASIAGTGAASNGYKGMAPGAKLAGIV 232

174 HPDLKSKVIGWDVAVNGRSTPYDDQGHVHAGIVAGTSV--NSQITGVAFGAKLVGRV 232

233 LGADSGSISTIIKGYEAVDNKDKYIKVINLSLSSQSSDGTSLSGAVNNAMDAIV 292

233 LGADSGSVSTIIAGVDMVQNKDKYIRIVNLSLSSQSSDGTSLSGAVNNAMDAIV 292

293 VCYAAGNSGPNTYVTSPPAASKVITVGAVDSDNINLASFSSRPTADGRLKEVYAPVD 352

293 VCYAAGNSGPNTYVTSPPAASKVITVGAVDSDNINLASFSSRPTADGRLKEVYAPVD 352

353 IIPRASGTSMTGPIIDYTTKASGTSMTATPHVSGVALILOAHPSTTPDKVKTALLETAD 412

353 IIPRASGTSMTGPIIDYTTKASGTSMTATPHVSGVALILOAHPSTTPDKVKTALLETAD 412

413 IYAPKEIADIAVAGRVNYKAIKYDVAKLFTGSGVADKGSATHTFDVSGATFVATLY 472

413 IYAPKEIADIAVAGRVNYKAIKYDVAKLFTGSGVADKGSATHTFDVSGATFVATLY 472

473 WDTGSSDIDLILYDPNGNEVDYSYTAIYGEKYGYNPNTAGTIVVYVSGKAAQVYV 532

473 WDTGSSDIDLILYDPNGNEVDYSYTAIYGEKYGYNPNTAGTIVVYVSGKAAQVYV 532

533 VSDGSLSGSGGPNPNPNPTPTDGTFTGSVNDYVDTSDTPTMNVNSGATKITGDL 592

Db 533 VSDGSLSGSGGPNPNPNPTPTDGTFTGSVNDYVDTSDTPTMNVNSGATKITGDL 592

QY 593 TFDTSYNDLILYDPNGNLVDRTSSNSYEHVEYANPAPGWTFLVYASTYGNADYOL 652

Db 593 TFDTSYNDLILYDPNGNLVDRTSSNSYEHVEYANPAPGWTFLVYASTYGNADYOL 652

QY 653 KAVVYVG 659

Db 653 KAVVYVG 659

RESULT 11

US-09-841-553-3

Sequence 3, Application US/09841553

Publication No. US20020086402A1

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru

APPLICANT: MORISHITA, Mio

APPLICANT: YAMAMOTO, Katsuhiko

APPLICANT: MITTA, Masanori

APPLICANT: ASADA, Kiyozo

APPLICANT: TSUNASAWA, Susumu

KATO, Ikunoshin

TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Broadway and Neimark

STREET: 419 Seventh Street N.W., Ste. 300

CITY: Washington

STATE: D.C.

COUNTRY: United States of America

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/841,553

FILING DATE: 24-Apr-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/894,818

FILING DATE: <Unknown>

APPLICATION NUMBER: JP 323285/1995

FILING DATE: 12-Dec-1995

ATTORNEY/AGENT INFORMATION:

NAME: Browdy, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TAKAKURA=1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 628-5197

TELEFAX: (202) 737-3528

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 522 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

OTHER INFORMATION:

Xaa at position 428 is Gly or Val.

SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-841-553-3

Query Match

Best Local Similarity 68.5%; Score 2349.5; DB 3; Length 522;

Matches 449; Conservative 22; Mismatches 51; Indels 5; Gaps 1;

1 ALEGLDESAQAQVMATYVWNLGYDSSGITIGIIDGIDASHPDLQGVYIGWVDFVNGRSY 192

1 ALEGLDESAQAQVMATYVWNLGYDSSGITIGIIDGIDASHPDLQGVYIGWVDFVNGRSY 192

```

Qy 193 PYDDHGHGTHVASIAAGTGAASNGKYKGMAPGAKLAGIVLQADSGSISTIIKGVEMAV 252
Db 61 PYDDHGHGTHVASIAAGTGAASNGKYKGMAPGAKLAGIVLQADSGSISTIIKGVEMAV 120
Qy 253 DNKDKYGIKVINLSLSSQSSDGTDSLGAQVANNAMDAIVVCVAAGNSGPNYTVGSPAA 312
Db 121 DNKDKYGIKVINLSLSSQSSDGTDSLGAQVANNAMDAIVVCVAAGNSGPNYTVGSPAA 180
Qy 313 ASKVIITVGAVDNDNIASFSSRGPTADGRLEKEVVAQGVDIAPRASGISMGTINDYTT 372
Db 181 ASKVIITVGAVDNDNIASFSSRGPTADGRLEKEVVAQGVDIAPRASGISMGTINDYTT 240
Qy 373 KASGTSMAATPHVSGVAGALLQAHPSWTPDKVKTALIEADIAVKEIADIAVAGRVNAY 432
Db 241 AAPTSMATPHVAGIALLQAHPSWTPDKVKTALIEADIAVKEIADIAVAGRVNAY 300
Qy 433 KAIKYDDYAKLFTFGSVADKSGATHTFDVSGATFTVATLTYMDTSSDIDLTYDPNGNEV 492
Db 301 KAIKYDDYAKLFTFGSVADKSGATHTFDVSGATFTVATLTYMDTSSDIDLTYDPNGNEV 360
Qy 493 DYSYTAAYGFEKVGYNPNPAGTWTVKVSYKGAANYQVDVSDGSLSSGSGGNPNPNP 552
Db 361 DYSYTAAYGFEKVGYNPNPAGTWTIKVSYSGSANYQVDVSDGSLSSGSGGNPNPNP 415
Qy 553 NPPTTDTGTFTGSVNDYMDTSDFTTMVNSGATKITGDLTFDTSYNDLTYLDPNGNL 612
Db 416 QPEPTVDAKTFQXSDHYHYDSDFTFTMTVNSGATKITGDLTFDTSYNDLTYLDPNGNL 475
Qy 613 VDRSSNSYEHVEYANAPAGTWTFLVYASTYGMADYQAKAVYYG 659
Db 476 VDRSSNSYEHVEYLTLPAGTWTFLVYASTYGMADYQAKAVYYG 522

```

```

RESULT 12
US-10-090-624-4
; Sequence 4, Application US/10090624
; Publication No. US2002012335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
; NAME/KEY: misc_feature
; LOCATION: (428)..(428)
; OTHER INFORMATION: Xaa at position 428 is Gly or Val.
US-10-090-624-4

```

```

Query Match 68.5%; Score 2349.5; DB 4; Length 522;
Best Local Similarity 85.2%; Pred. No. 9e-158;
Matches 449; Conservative 22; Mismatches 51; Indels 5; Gaps 1;
Qy 133 AELBGLDESAQVMATYVMNLGYDGSIGITIGIITDGDASHPDLOGKVIQWDFVNGRSY 192
Db 1 AELBGLDESAQVMATYVMNLGYDGSIGITIGIITDGDASHPDLOGKVIQWDFVNGRSY 60
Qy 193 PYDDHGHGTHVASIAAGTGAASNGKYKGMAPGAKLAGIVLQADSGSISTIIKGVEMAV 252

```

```

Db 61 PYDDHGHGTHVASIAAGTGAASNGKYKGMAPGAKLAGIVLQADSGSISTIIKGVEMAV 120
Qy 253 DNKDKYGIKVINLSLSSQSSDGTDSLGAQVANNAMDAIVVCVAAGNSGPNYTVGSPAA 312
Db 121 DNKDKYGIKVINLSLSSQSSDGTDSLGAQVANNAMDAIVVCVAAGNSGPNYTVGSPAA 180
Qy 313 ASKVIITVGAVDNDNIASFSSRGPTADGRLEKEVVAQGVDIAPRASGISMGTINDYTT 372
Db 181 ASKVIITVGAVDNDNIASFSSRGPTADGRLEKEVVAQGVDIAPRASGISMGTINDYTT 240
Qy 373 KASGTSMAATPHVSGVAGALLQAHPSWTPDKVKTALIEADIAVKEIADIAVAGRVNAY 432
Db 241 AAPTSMATPHVAGIALLQAHPSWTPDKVKTALIEADIAVKEIADIAVAGRVNAY 300
Qy 433 KAIKYDDYAKLFTFGSVADKSGATHTFDVSGATFTVATLTYMDTSSDIDLTYDPNGNEV 492
Db 301 KAIKYDDYAKLFTFGSVADKSGATHTFDVSGATFTVATLTYMDTSSDIDLTYDPNGNEV 360
Qy 493 DYSYTAAYGFEKVGYNPNPAGTWTVKVSYKGAANYQVDVSDGSLSSGSGGNPNPNP 552
Db 361 DYSYTAAYGFEKVGYNPNPAGTWTIKVSYSGSANYQVDVSDGSLSSGSGGNPNPNP 415
Qy 553 NPPTTDTGTFTGSVNDYMDTSDFTTMVNSGATKITGDLTFDTSYNDLTYLDPNGNL 612
Db 416 QPEPTVDAKTFQXSDHYHYDSDFTFTMTVNSGATKITGDLTFDTSYNDLTYLDPNGNL 475
Qy 613 VDRSSNSYEHVEYANAPAGTWTFLVYASTYGMADYQAKAVYYG 659
Db 476 VDRSSNSYEHVEYLTLPAGTWTFLVYASTYGMADYQAKAVYYG 522

```

```

RESULT 13
US-10-800-684-3
; Sequence 3, Application US/10800684
; Publication No. US2005001422A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: ULTRATHERMOSTABLE PROTEASE GENES
; FILE REFERENCE: TAKAKURA-1A
; CURRENT APPLICATION NUMBER: US/10/800,684
; PRIOR FILING DATE: 2004-03-16
; PRIOR APPLICATION NUMBER: US/09/841,553
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 08/894,818
; PRIOR FILING DATE: 1997-08-29
; PRIOR APPLICATION NUMBER: JP2385/1995
; PRIOR FILING DATE: 1995-12-12
; PRIOR APPLICATION NUMBER: JP96/03253
; PRIOR FILING DATE: 1996-11-07
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
; NAME/KEY: misc_feature
; LOCATION: (428)..(428)
; OTHER INFORMATION: Xaa is Gly or Val
US-10-800-684-3

```

```

Query Match 68.5%; Score 2349.5; DB 5; Length 522;
Best Local Similarity 85.2%; Pred. No. 9e-158;
Matches 449; Conservative 22; Mismatches 51; Indels 5; Gaps 1;
Qy 133 AELBGLDESAQVMATYVMNLGYDGSIGITIGIITDGDASHPDLOGKVIQWDFVNGRSY 192

```

```
DB 1 ALEGLDESAQAQVMAVYWNVLGYDGSGITIGIIDTGIDASHPDLQCKVIGWDFVNGRSY 60
QY 193 PYDDHGHGTHVASIAAGTGAASNGKYGMAPKAGLAGIKVIGADSGSISTIIKGVEMAV 252
DB 61 PYDDHGHGTHVASIAAGTGAASNGKYGMAPKAGLAGIKVIGADSGSISTIIKGVEMAV 120
QY 253 DNKDKYGIKVINLSLSSQSSDGTDLAQVANNAMDAIIVCAAGNSGPNITYVGSPPA 312
DB 121 DNKDKYGIKVINLSLSSQSSDGTDLAQVANNAMDAIIVCAAGNSGPNITYVGSPPA 180
QY 313 ASKVTIVGAVDNDNIASFSSRGPTADGRLKEPVAVPGVDIIAPRAGSTSGTPIINDYTT 372
DB 181 ASKVTIVGAVDNDNIASFSSRGPTADGRLKEPVAVPGVDIIAPRAGSTSGTPIINDYTT 240
QY 373 KASGSMATPHVSGVALLILOAHPSMTDPKXTALIEADIAPEIADIAAGRVAVY 432
DB 241 AAPGTSMAITPHVAGIALLILOAHPSMTDPKXTALIEADIAPEIADIAAGRVAVY 300
QY 433 KAIKTDYAKLTFTGSDVADKGSATHTFDVSGATFVTATLYMDTSSDIDLVLYPDNGNEV 492
DB 301 KAINDNVAKLVFTGYVANKSGQTHQFVLSGASFTVATLYMDNANSDLDLVLYPDNGNOV 360
QY 493 DYSYTAAYGFEKVGYYNPTAGTWTIVKVSYKGAANYQVDVSDGSLSSQSGGPNPNPNP 552
DB 361 DYSYTAAYGFEKVGYYNPTAGTWTIVKVSYKGAANYQVDVSDGSLSSQSGGPNPNPNP 415
QY 553 NPTPTDQTFPGSVNDWDSPTFTMNVNSGATKITGDLTPTDTSYNDLDLVLYPDNGNL 612
DB 416 QPEPVDATFOXSDHYDRSDPTFTMNVNSGATKITGDLTPTDTSYNDLDLVLYPDNGNL 475
QY 613 VDRSTSSNHYEVEYANPAGTWTFLVYASTYGMADYOLKAVVYVG 659
DB 476 VDRSTSSNHYEVEYANPAGTWTFLVYASTYGMADYOLKAVVYVG 522
```

```
RESULT 14
US-10-888-588-4
; Sequence 4, Application US/10888588
; Publication No. US20050084934A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hi-karu
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOTO, Tomoko
; APPLICANT: ASADA, Kiyo-ozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/888,588
; CURRENT FILING DATE: 2004-07-12
; PRIOR APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
; FEATURE:
; NAME/KEY: misc. feature
; LOCATION: (428)..(428)
; OTHER INFORMATION: Xaa at position 428 is Gly or Val.
US-10-888-588-4
```

```
Query Match 68.5%; Score 2349.5; DB 5; Length 522;
Best Local Similarity 85.2%; Pred. No. 9e-158;
Matches 449; Conservative 22; Mismatches 51; Indels 5; Gaps 1;
QY 133 ALEGLDESAQAQVMAVYWNVLGYDGSGITIGIIDTGIDASHPDLQCKVIGWDFVNGRSY 192
```

```
DB 1 ALEGLDESAQAQVMAVYWNVLGYDGSGITIGIIDTGIDASHPDLQCKVIGWDFVNGRSY 60
QY 193 PYDDHGHGTHVASIAAGTGAASNGKYGMAPKAGLAGIKVIGADSGSISTIIKGVEMAV 252
DB 61 PYDDHGHGTHVASIAAGTGAASNGKYGMAPKAGLAGIKVIGADSGSISTIIKGVEMAV 120
QY 253 DNKDKYGIKVINLSLSSQSSDGTDLAQVANNAMDAIIVCAAGNSGPNITYVGSPPA 312
DB 121 DNKDKYGIKVINLSLSSQSSDGTDLAQVANNAMDAIIVCAAGNSGPNITYVGSPPA 180
QY 313 ASKVTIVGAVDNDNIASFSSRGPTADGRLKEPVAVPGVDIIAPRAGSTSGTPIINDYTT 372
DB 181 ASKVTIVGAVDNDNIASFSSRGPTADGRLKEPVAVPGVDIIAPRAGSTSGTPIINDYTT 240
QY 373 KASGSMATPHVSGVALLILOAHPSMTDPKXTALIEADIAPEIADIAAGRVAVY 432
DB 241 AAPGTSMAITPHVAGIALLILOAHPSMTDPKXTALIEADIAPEIADIAAGRVAVY 300
QY 433 KAIKTDYAKLTFTGSDVADKGSATHTFDVSGATFVTATLYMDTSSDIDLVLYPDNGNEV 492
DB 301 KAINDNVAKLVFTGYVANKSGQTHQFVLSGASFTVATLYMDNANSDLDLVLYPDNGNOV 360
QY 493 DYSYTAAYGFEKVGYYNPTAGTWTIVKVSYKGAANYQVDVSDGSLSSQSGGPNPNPNP 552
DB 361 DYSYTAAYGFEKVGYYNPTAGTWTIVKVSYKGAANYQVDVSDGSLSSQSGGPNPNPNP 415
QY 553 NPTPTDQTFPGSVNDWDSPTFTMNVNSGATKITGDLTPTDTSYNDLDLVLYPDNGNL 612
DB 416 QPEPVDATFOXSDHYDRSDPTFTMNVNSGATKITGDLTPTDTSYNDLDLVLYPDNGNL 475
QY 613 VDRSTSSNHYEVEYANPAGTWTFLVYASTYGMADYOLKAVVYVG 659
DB 476 VDRSTSSNHYEVEYANPAGTWTFLVYASTYGMADYOLKAVVYVG 522
```

```
RESULT 15
US-10-090-624-1
; Sequence 1, Application US/10090624
; Publication No. US2002013235A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hi-karu
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOTO, Tomoko
; APPLICANT: ASADA, Kiyo-ozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-10-090-624-1
```

```
Query Match 56.0%; Score 1918; DB 4; Length 412;
Best Local Similarity 89.5%; Pred. No. 2.4e-127;
Matches 367; Conservative 15; Mismatches 28; Indels 0; Gaps 0;
QY 133 ALEGLDESAQAQVMAVYWNVLGYDGSGITIGIIDTGIDASHPDLQCKVIGWDFVNGRSY 192
DB 1 ALEGLDESAQAQVMAVYWNVLGYDGSGITIGIIDTGIDASHPDLQCKVIGWDFVNGRSY 60
QY 193 PYDDHGHGTHVASIAAGTGAASNGKYGMAPKAGLAGIKVIGADSGSISTIIKGVEMAV 252
DB 61 PYDDHGHGTHVASIAAGTGAASNGKYGMAPKAGLAGIKVIGADSGSISTIIKGVEMAV 120
```

```

QY 253 DNRKXGIXVNLISLSSQSSDGTSLSOAVNNAMDAGIVVCAAGNSGPNITYTVGSPAA 312
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 DNRKXGIXVNLISLSSQSSDGTSLSOAVNNAMDAGLVVVVAAGNSGPNKTYTIGSPAA 180
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 313 ASKVIITVGAVDSNDNTIASFSSRGPTADGRKPEVVAPEGVDIIAPRASGTMGTPINDYYT 372
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 ASKVIITVGAVDKYDVITSFSSRGPTADGRKPEVVAPEGVDIIAPRASGTMGTPINDYYT 240
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 373 KASGTSNATPHVSGVGLIIOAHPSWTPDKVKTALITETADIIVAPKEIADIAYGAGRNVY 432
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 241 AAPGTSNATPHVAGIALLIQAHPSWTPDKVKTALITETADIIVAPKEIADIAYGAGRNVY 300
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 433 KAIKYDDYAKLFTGVSADKGSATHTFVDSGATFVTATLYMDTGSSPDLDLYLDPNGNEV 492
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 301 KAINYDNYAKLVFTGVANKGSOHQHGFVISGASFVTATLYMDNANSDDLXYLDPNGNIV 360
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 493 DYSYTAIYGFGEKGYNPTAGTWTVKVVSYKGAANTQVVDVSDGSLSQSG 542
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 361 DYSYTAIYGFGEKGYNPTDGTWTIKVVSYSGSANYQVDVSDGSLSQSG 410
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

Search completed: January 6, 2007, 22:37:51  
 Job time : 125.5 secs



GenCore/version 5.1.9  
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2007, 22:33:57 ; Search time 31 Seconds  
(without alignments)  
2087.781 Million cell updates/sec

Title: US-10-800-684-5  
Perfect score: 3428  
Sequence: 1 MKGLKALIVLVLGLVGS.....YAVSTYGMADYQLKAVVYVG 659

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 393712 seqs, 98211237 residues  
Total number of hits satisfying chosen parameters: 393712

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications\_AA\_New:\*

- 1: /EMC\_Celerra\_SIDS3/Ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*
- 2: /EMC\_Celerra\_SIDS3/Ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*
- 3: /EMC\_Celerra\_SIDS3/Ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*
- 4: /EMC\_Celerra\_SIDS3/Ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*
- 5: /EMC\_Celerra\_SIDS3/Ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*
- 6: /EMC\_Celerra\_SIDS3/Ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*
- 7: /EMC\_Celerra\_SIDS3/Ptodata/1/pubpaa/US11\_NEW\_PUB pep.\*
- 8: /EMC\_Celerra\_SIDS3/Ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                           |
|------------|-------|-------------|--------|-------|---------------------------------------|
| 1          | 2705  | 78.9        | 663    | 6     | US-10-526-324-251 Sequence 251, App   |
| 2          | 670   | 19.5        | 443    | 6     | US-10-530-643-8 Sequence 8, Appl      |
| 3          | 660   | 19.3        | 1227   | 6     | US-10-530-643-6 Sequence 6, Appl      |
| 4          | 588   | 17.2        | 552    | 6     | US-10-530-643-124 Sequence 124, App   |
| 5          | 557   | 16.2        | 379    | 7     | US-11-433-614-14 Sequence 14, Appl    |
| 6          | 539   | 15.7        | 1606   | 6     | US-10-530-643-114 Sequence 114, App   |
| 7          | 532   | 15.5        | 274    | 6     | US-10-532-605-2 Sequence 2, Appl      |
| 8          | 527   | 15.4        | 381    | 7     | US-11-452-695-2 Sequence 2, Appl      |
| 9          | 525   | 15.3        | 274    | 6     | US-10-541-737-3 Sequence 3, Appl      |
| 10         | 524   | 15.3        | 381    | 7     | US-11-452-695-3 Sequence 3, Appl      |
| 11         | 524   | 15.3        | 381    | 7     | US-11-452-695-4 Sequence 4, Appl      |
| 12         | 518   | 15.1        | 381    | 7     | US-11-433-614-16 Sequence 16, Appl    |
| 13         | 517.5 | 15.1        | 275    | 6     | US-10-567-073-3 Sequence 3, Appl      |
| 14         | 517.5 | 15.1        | 275    | 6     | US-10-541-737-1 Sequence 1, Appl      |
| 15         | 504.5 | 14.7        | 524    | 6     | US-10-526-324-20 Sequence 20, Appl    |
| 16         | 501.5 | 14.6        | 275    | 6     | US-10-541-737-2 Sequence 2, Appl      |
| 17         | 494.5 | 14.4        | 624    | 6     | US-10-530-643-46 Sequence 46, Appl    |
| 18         | 493.5 | 14.4        | 269    | 6     | US-10-541-737-4 Sequence 4, Appl      |
| 19         | 491   | 14.3        | 274    | 7     | US-11-452-695-1 Sequence 1, Appl      |
| 20         | 486.5 | 14.2        | 673    | 6     | US-10-530-643-56 Sequence 56, Appl    |
| 21         | 481   | 14.0        | 422    | 6     | US-10-526-324-979 Sequence 979, App   |
| 22         | 458.5 | 13.4        | 462    | 6     | US-10-449-902-40515 Sequence 40515, A |
| 23         | 456   | 13.3        | 400    | 6     | US-10-530-643-60 Sequence 60, Appl    |
| 24         | 454.5 | 13.3        | 469    | 6     | US-10-449-902-37064 Sequence 37064, A |
| 25         | 449   | 13.1        | 518    | 6     | US-10-530-643-122 Sequence 122, App   |

|    |       |      |     |   |                                       |
|----|-------|------|-----|---|---------------------------------------|
| 26 | 436.5 | 12.7 | 269 | 6 | US-10-516-164A-1 Sequence 1, Appl     |
| 27 | 432.5 | 12.6 | 607 | 6 | US-10-530-643-62 Sequence 62, Appl    |
| 28 | 428.5 | 12.5 | 615 | 6 | US-10-530-643-88 Sequence 88, Appl    |
| 29 | 426.5 | 12.4 | 523 | 6 | US-10-530-643-40 Sequence 4, Appl     |
| 30 | 424.5 | 12.4 | 640 | 7 | US-11-318-576-4 Sequence 4, Appl      |
| 31 | 415   | 12.1 | 611 | 6 | US-10-530-643-74 Sequence 74, Appl    |
| 32 | 412   | 12.0 | 434 | 7 | US-11-318-576-10 Sequence 10, Appl    |
| 33 | 403.5 | 11.8 | 434 | 6 | US-10-530-643-128 Sequence 128, App   |
| 34 | 403   | 11.8 | 434 | 6 | US-11-318-576-15 Sequence 15, Appl    |
| 35 | 402   | 11.7 | 434 | 6 | US-10-530-643-2 Sequence 2, Appl      |
| 36 | 401.5 | 11.7 | 433 | 7 | US-11-318-576-12 Sequence 12, Appl    |
| 37 | 401.5 | 11.7 | 433 | 7 | US-11-318-576-14 Sequence 14, Appl    |
| 38 | 401   | 11.7 | 434 | 7 | US-11-318-576-2 Sequence 2, Appl      |
| 39 | 401   | 11.7 | 434 | 7 | US-11-318-576-11 Sequence 11, Appl    |
| 40 | 395.5 | 11.5 | 433 | 7 | US-11-318-576-13 Sequence 13, Appl    |
| 41 | 391.5 | 11.4 | 515 | 6 | US-10-530-643-50 Sequence 50, Appl    |
| 42 | 390   | 11.4 | 433 | 7 | US-11-318-576-16 Sequence 16, Appl    |
| 43 | 389.5 | 11.4 | 759 | 6 | US-10-449-902-47258 Sequence 47258, A |
| 44 | 387   | 11.3 | 754 | 6 | US-10-449-902-47240 Sequence 47240, A |
| 45 | 381.5 | 11.1 | 426 | 6 | US-10-530-643-42 Sequence 42, Appl    |

#### ALIGNMENTS

RESULT 1  
US-10-526-324-251  
Sequence 251, Application US/10526324  
Publication No. US20060248617A1  
GENERAL INFORMATION:  
APPLICANT: Imanaka, Takayuki  
APPLICANT: Atom, Haruyuki  
TITLE OF INVENTION: METHOD OF TARGETED GENE DISRUPTION, GENOME OF  
TITLE OF INVENTION: HYPERHERMOSTABLE BACTERIUM AND GENOME CHIP USING  
TITLE OF INVENTION: THE SAME  
FILE REFERENCE: 490051.401USPC  
CURRENT APPLICATION NUMBER: US/10/526,324  
CURRENT FILING DATE: 2005-02-28  
PRIOR APPLICATION NUMBER: PCT/IB2003/003597  
PRIOR FILING DATE: 2003-08-29  
PRIOR APPLICATION NUMBER: JP 2002-319011  
PRIOR FILING DATE: 2002-08-30  
NUMBER OF SEQ ID NOS: 2167  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 251  
LENGTH: 663  
TYPE: PRT  
ORGANISM: Thermococcus kodakaraensis KOD1  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (414562)..(414562)  
OTHER INFORMATION: n is a or c or g or t.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (786890)..(786890)  
OTHER INFORMATION: n is a or c or g or t.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (786890)..(786890)  
OTHER INFORMATION: n is a or c or g or t.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (786907)..(786907)  
OTHER INFORMATION: n is a or c or g or t.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (839139)..(839139)  
OTHER INFORMATION: n is a or c or g or t.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1128488)..(1128488)  
OTHER INFORMATION: n is a or c or g or t.  
FEATURE:

```

NAME/KEY: misc feature
LOCATION: (1128499)..(1128499)
OTHER INFORMATION: n is a or c or g or t.
FEATURE:
NAME/KEY: misc feature
LOCATION: (1128505)..(1128506)
OTHER INFORMATION: n is a or c or g or t.
FEATURE:
NAME/KEY: misc feature
LOCATION: (1128517)..(1128518)
OTHER INFORMATION: n is a or c or g or t.
FEATURE:
NAME/KEY: misc feature
LOCATION: (1128539)..(1128540)
OTHER INFORMATION: n is a or c or g or t.
FEATURE:
NAME/KEY: misc feature
LOCATION: (1561400)..(1561400)
OTHER INFORMATION: n is a or c or g or t.
FEATURE:
NAME/KEY: misc feature
LOCATION: (1561477)..(1561477)
OTHER INFORMATION: n is a or c or g or t.
FEATURE:
NAME/KEY: misc feature
LOCATION: (1561545)..(1561545)
OTHER INFORMATION: n is a or c or g or t.
FEATURE:
NAME/KEY: misc feature
LOCATION: (1767941)..(1767941)
OTHER INFORMATION: n is a or c or g or t.
US-10-526-324-251

```

```

Query Match
Best Local Similarity 78.9%; Score 2705; DB 6; Length 663;
Matches 517; Conservative 60; Mismatches 80; Indels 8; Gaps 4;

```

```

QY 1 MKGKALILVLVGLVGVSAAPBEKKEVQYRV--EKYGLITPGLFKIQKLPNEP 58
DB 1 MKKGAVALLFLVGLMAGSVLAAPQKPA--VRNVSOQKXGGLTLPGLFKVQRMWDCE 58
QY 59 ISTVIVFNHREKEIAVRVLEIMGAKRYVYHIIIPALADLKVRDLVLVIGL--TG--GK 114
DB 59 VSTIMFDNQADKKKAVIIDFLGAKIKNYHIIIPALAVIKYKDLIINGLMDTGYFGN 118
QY 115 AKLSGVRFIQEDYKVTYSAIEGLDESAAQVMATYVNIIGYDGGITIGIIDGIDASHP 174
DB 119 AQLSGVQFIQEDYVYKVAVELEGIDBSAAQVMATNMNIGYDGGITIGIIDGIDASHP 178
QY 175 DLQGVVGVWDFVNGRGEPYDDHGHGTHVASIAGCGAANGKYKGMAPAKLAGIKVLG 234
DB 179 DLQGVVGVWDFVNGKTPYDDHGHGTHVASIAGCGAANGKYKGMAPAKLAGIKVLN 238
QY 235 ADGSGSISTIIKGVEMAVDNKKYKIKVINLSGSSQSSDGTDSLQAVNANMADAGIYVC 294
DB 239 GQSGSISDIINGVDMAVQKDKKIKVINLSGSSQSSDGTDSLQAVNANMADAGIYV 298
QY 295 VAANGSGPNTYTGSPPAASKVITVGAVDNNDIASFSSRGPTADGRLKEVVAPEVDII 354
DB 299 VAANGSGPNTYTGSPPAASKVITVGAVDYITTFSSRGPTADNRLKEVVAPEVVAPEVDII 358
QY 355 APASGTSMTPTINDYTTKASGTSMATPHYSGVGLIIOAHPEWTDKXTALIEFADIV 414
DB 359 AARASGTSMTPTINDYTTAAPTSMATPHYAGIAALLQAHPEWTDKXTALIEFADIV 418
QY 415 APEIADIAVAGRVVYKAIKYDDYAKLFTGSVADKGSATTFPVSAGTEVATLYMD 474
DB 419 KPBEIADIAVAGRVVYKAAAYYDNYAKLFTGYVSNKGSQSHQFTISGAGFATLYMD 478
QY 475 TGSSDIDLIVYDNGENVDYSYTAAYYGFEKVGYNPTAGTWTYKVSYYGAANYQYDVVS 534
DB 479 NGSDDLIVYDNGENVDYSYTAAYYGFEKVGYNPTAGTWTYKVSYYGSANYQYDVVS 538

```

```

QY 535 DGSLSQSGGNPNPNPNPTPTDPTQTFGSVNDYWDTSDTFTMNVNSGATKITGDLTF 594
DB 539 DGSLSQSGGSGSPSPSPVDEKTFGTGYHDYDKSDTFTMNVNSGATKITGDLTF 598
QY 595 DTSYNDIDLIVDPNGNLVDRSTSSNSYEHVEYANPAGTWTLYVYASTYGNADYQJKA 654
DB 599 DTSYNDIDLIVDPNGNLVDRSSSSNSYEHVEYANNPAGTWTLYVYADTYGADYQJLDA 658
QY 655 VVYVG 659
DB 659 KVIYG 663

```

## RESULT 2

```

US-10-530-643-8
Sequence 8, Application US/10530643
Publication No. US2006025995A1
GENERAL INFORMATION:
APPLICANT: CAYOURETTE, Michelle
APPLICANT: HANSEN, Connie Jo
APPLICANT: MCCLEURE, Amy
APPLICANT: SUN, May
APPLICANT: GRAMATIKOVA, Svetlana
APPLICANT: DYCAICO, Mark
APPLICANT: BARTON, Nelson R.
APPLICANT: STEGE, Justin T.
APPLICANT: ABOUSHADI, Nahla M.
TITLE OF INVENTION: PROTEASES, NUCLEIC ACIDS ENCODING THEM
FILE REFERENCE: 564462004100
CURRENT APPLICATION NUMBER: US/10/530,643
PRIOR FILING DATE: 2006-04-07
PRIOR APPLICATION NUMBER: PCT/US03/32819
PRIOR FILING DATE: 2003-10-10
PRIOR APPLICATION NUMBER: 60/471,423
PRIOR FILING DATE: 2003-05-16
PRIOR APPLICATION NUMBER: 60/418,467
PRIOR FILING DATE: 2002-10-10
NUMBER OF SEQ ID NOS: 255
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 443
TYPE: PRT
ORGANISM: Bacteria
US-10-530-643-8

```

```

Query Match
Best Local Similarity 19.5%; Score 670; DB 6; Length 443;
Matches 161; Conservative 81; Mismatches 149; Indels 42; Gaps 11;

```

```

QY 30 EGVNRVKNYGLITPGLFKIQKLPN--EISTVIVFNHREKEIAVRVLEIM---GAK 84
DB 23 DRLNLVYKPFKMTPEFLHMFPERFKNRKKLSVILEFSSNCOBCELVNQAFTGRNCK 82
QY 85 VAYVYHIIIPALADLKVRDLVLVIGLTKGAKLVSVRIQEDYKVTYSAIEGLDES-- 142
DB 83 INNEFSIISCSADITPVL-----BEVLTWCNHIK--KYLLNHEVRALDLTAIVI 130
QY 143 -AQWATYVNMNGYVSGSITIGIIDGIDAGSPDLQGVVGVWDFVNGRGEPYDDHGHGT 201
DB 131 SANAKIYRANTYTLTGKITTAVIDTGL--FPHITLSGIIIFVDFINDIRIETTYDNGHGT 189
QY 202 HVAISIACTGAASNGKYKGMAPGAKLAGIKVYLGADGSGSISTIIKGVEMAV---DNKDKY 258
DB 190 HCAQDALDGDGASGGLYMGPAPEANVIGVTLNKKVSGSGLTVMQGVDMCIRYDQNDQ 249
QY 259 GIKVINLSIGS---SQSDGTDLSQAVNMAWDAGIYVCAAGNSGPTTYTGSPPAASK 315
DB 250 KINIINSLGAPAPQRYENENDDPVYKMYEKAMENGIYVCAAGNGBEPASTIASPVSQ 309
QY 316 VITVGAVD-----SNDNIASFSSRGPTADGRLKEVVAPEVVDIAPRASGTSM---- 363
DB 310 VITVGAVDKTTATTRBDBDEVASFSSRGPTIYKVKYDIIAPGVDIISLRSPNSYLDKQ 369

```



RESULT 5  
US-11-433-614-14  
; Sequence 14, Application US/11433614  
; Publication No. US20060205019A1  
; GENERAL INFORMATION:  
; APPLICANT: Athena Biotechnologies, Inc.  
; APPLICANT: Menzel, Rolf  
; TITLE OF INVENTION: Methods and Compositions for Directed Gene Assembly  
; FILE REFERENCE: 46675-5004-01-US  
; CURRENT APPLICATION NUMBER: US/11/433,614  
; CURRENT FILING DATE: 2006-05-12  
; PRIOR APPLICATION NUMBER: US 09/920,118  
; PRIOR FILING DATE: 2001-07-31  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 14  
; LENGTH: 379  
; TYPE: PRF  
; ORGANISM: Bacillus licheniformis  
US-11-433-614-14

Query Match 16.2%; Score 557; DB 7; Length 379;  
Best Local Similarity 34.3%; Pred. No. 1,8e-31;  
Matches 150; Conservative 64; Mismatches 149; Indels 74; Gaps 14;

QY 3 GRLKLLIVLVGLVGSVAAAPKRYBQVRNVEKNTGLTLPGLFRKIQKLNPNRISTV 62  
DB 10 GMLTALMLVFTMAPSDSASAAQPAK-----NVEKDY-----I 41  
QY 63 IVEFNH-REKEIAVRVLELMGAKYRVYVHIIIPALADLKVRLDVLVLSGLTGKRAKLSGR 121  
DB 42 VGFSGVKTASVKDDIIESGSKVDKQFRIINAKALDKBALREYV-----NDPVA 94  
QY 122 FIOGDYVTVSABLEGLDESAQAQVATYVNLGYDGSGITIGIIDTGIDASHPDLOQVY 181  
DB 95 YVEBDH--VAHALAQTPYGPILKADKVOAQGKGNVAVVLDITQASHDL--NVV 150  
QY 182 GWVDFVNGRYPVDDHGHGHTVHVASIAGTGAASNGK--YKGAAPGAKLAGIKVLGADSG 239  
DB 151 GGAAPVAGAEVNTDGNHGHV---AGTVAALDNTTGVGVAPNVLVAVKVLNASSGSG 206  
QY 240 SISTIINGEMAVDNKDKYKIVNLISLSSQSSDDTDSLSQAVNNAMDAIGIVVCAAGN 299  
DB 207 SYSGIVSGIEMATTN---GMDVIINSLG---GPSSTAMKQAVDANVARGVAVVAAAGN 259  
QY 300 SGP--NTYTGSPAAASKVITVGAVDSDNINASFSSRGPTADGRLEPEVVAPEGVDIIAPR 357  
DB 260 SGSSGNTNTIGYPAKYDSVIAVAVDPNSNRASFSSVG-----AELEVAAPGAGVYSTY 313  
QY 358 ASGTSMGTPIINDYTTASGSMATPHVSGVGLILQAHPSMTDKYKTALILETADIVAKR 417  
DB 314 PTGT-----YATLNGTSMASPHVAGAAALILSKHPVLASQVRNRLSTATYLG-- 362  
QY 418 EIADIYAGAGRVNVYKA 434  
DB 363 --SSFYKGLINVEAA 377

RESULT 6  
US-10-530-643-114  
; Sequence 114, Application US/10530643  
; Publication No. US2006025995A1  
; GENERAL INFORMATION:  
; APPLICANT: CAYOUEITE, Michelle  
; APPLICANT: HANSEN, Connie Jo  
; APPLICANT: MCCLORE, Amy  
; APPLICANT: SUN, May  
; APPLICANT: GRAMATIKOVA, Svetlana  
; APPLICANT: DYCAICO, Mark  
; APPLICANT: BARTON, Nelson R.  
; APPLICANT: STEGE, Justin T.  
; APPLICANT: ABOUSHADI, Nahla M.  
; TITLE OF INVENTION: PROTEASES, NUCLEIC ACIDS ENCODING THEM

; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM  
; FILE REFERENCE: 564462004100  
; CURRENT APPLICATION NUMBER: US/10/530,643  
; CURRENT FILING DATE: 2006-04-07  
; PRIOR APPLICATION NUMBER: PCT/US03/32819  
; PRIOR FILING DATE: 2003-10-10  
; PRIOR APPLICATION NUMBER: 60/471,423  
; PRIOR FILING DATE: 2003-05-16  
; PRIOR APPLICATION NUMBER: 60/418,467  
; PRIOR FILING DATE: 2002-10-10  
; NUMBER OF SEQ ID NOS: 255  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 114  
; LENGTH: 1606  
; TYPE: PRF  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample.  
; NAME/KEY: SIGNAL  
; LOCATION: (1)...(32)  
US-10-530-643-114

Query Match 15.7%; Score 539; DB 6; Length 1606;  
Best Local Similarity 22.6%; Pred. No. 2.7e-29;  
Matches 233; Conservative 111; Mismatches 235; Indels 454; Gaps 37;

QY 2 KGLKLLIVLVGLVGSVAAAPKRYBQVRNVEKNTGLTLPGLFRKIQKLNPNRISTV 45  
DB 8 KFLSAFLTILVLSLVPNGTANALAKQPKTNPQDASALQQAALAQOTSLLSGAR 67  
QY 46 LFRKIQKLNPNRISTV-----VFNHREK 71  
DB 68 LHKDQLSGSGQEVVILHLSKAVGLBOGHLKNGKMSQSEITLTKGKTQAOQNTARK 127  
QY 72 ELAVRVLELMGAKYRV--YTH---IIPALADLKVRLDVLVLSGLTGKRAKLSGRFIGED 126  
DB 128 EMTVK-----KVFEDVYSYDTVLNGLSGTVQANDLKL--LT-----ISGVYVSPD 173  
QY 127 YKVTYSABLEGLDESAQAQVATYV-----VNLGYDGSGITIGIIDTGIDASHPDLOQ 178  
DB 174 TTVYAOBGLQVDPKYVDKAMDTISPLFQINKLMEBGIQGVKAVVLDITGIDASHPDKA 233  
QY 179 KVIQWDFVNGRYPYV---DDHG-----HGTVAASIAA 208  
DB 224 YKGGKNTFVPHVDGYAPRADDSSTSPDRPANKREVENRGSSFTTGTHTVAGTIV 293  
QY 209 GTGAASNGKTYGMAPGAKLAGIKVLGADSGSISTITIKGEMAVDNKDKYKIVNLISLG 268  
DB 294 AQGNNEFG-IRGIAEKVDLYSYRVLGAYSGATSGIIRKIDTAVIEK---IDVINLSLG 348  
QY 269 SSQSSDGTDSLSQAVNNAMDAIGIVVCAAGNSGPTVVGSPAAASKYITVG----- 320  
DB 349 GGANSE-IDGASPAINNAMLAGTISVATGNSGRNGTGTIPATISRLGIANGTTPETH 407  
QY 321 ----- 320  
DB 408 FOANVNVAGEFTYSKVNDLMAITFGKOVETOLAGYDLIAPGAGAKOYAADVYTGKV 467  
QY 321 -----AVD----- 323  
DB 468 ALVABSEIAPVDKVAALKKAGAVAIIVHNFAGGTNAENKSDVFLGDAFEFIPFDMGVTD 527  
QY 324 -----SNDNIASFSSRGPTADGRLEPEVVAPEGVDIIAPR 355  
DB 528 GEALRAALAAPSTISPDQFSIKITGDDVDVNDSSRSPSTNPFIKPDVABPGINIMSTI 587  
QY 356 PRASGTSMGTPIINDYTTASGSMATPHVSGVGLILQAHPSMTDKYKTALILETADIVA 415  
DB 588 PMYNDPPDADYSBAYSRKGTSMATPHIAGIALVAVQAMPVNAFVVKVVALSNTATVLD 647  
QY 416 PEIADIYAGAGRVNVKAIKYD----- 438

Db 648 TKKYDVFAQAGRVADYKARADVLAYALDTASNDGTEVENIKGVTTFGPQKLDKNISVT 707  
 QY 439 -----DVA-----XLTFTGSVAD----- 451  
 Db 708 KKNVVDLKSAGDVTYVSDVTKGPDGAKVTVDQSEFTLNGEQLLVNTLTASAEETKAGD 767  
 QY 452 -----KG-----SATHFDVSGATFVT-----ATLYW----- 473  
 Db 768 ELIYGTHIKNGKMDLSLPFAADGGAAYAVKDMETKTDLSPNGDVNDMLLYFTLIG 827  
 QY 474 DPGSSIDLY-LYDPNGNEVDYSYTAHYGPEKGYTNP-----TAGWTW-----KVSYSYG 524  
 Db 828 DVGFNTIELMDIMNPTGKXSDY-----IGYHASDPTLTAGSYOLPLLEKTYOWOG 879  
 QY 525 AANYQDVVSDGSL-----SOSGGGNPNPNP-NPPTPTDQ-TFTGSVND----- 569  
 Db 880 TMS---EVLPGDLTYTIDFSAETSGNPKT1ADYGVVVVXKSTAGTLEGAVADGKTGKIT 936  
 QY 570 ---YMD-----TSPTFTMNVN-----SGATKITGD-----LTPTSY 598  
 Db 937 DKYVDQKELVKYKMGVYDNTKLSATYEVYANDAVVDSGAVKLAQDGTAFDLPTFDKX 996  
 QY 599 NDLDLYLYDPNGN 611  
 Db 997 NNVTYKXADAGN 1009

RESULT 7  
 US-10-532-605-2  
 ; Sequence 2, Application US/10532605  
 ; Publication No. US20060134092A1

; GENERAL INFORMATION:  
 ; APPLICANT: Miwa, Takehiro  
 ; TITLE OF INVENTION: Method for digesting proteins highly resistant to denaturation  
 ; FILE REFERENCE: Q87625  
 ; CURRENT APPLICATION NUMBER: US/10/532,605  
 ; CURRENT FILING DATE: 2005-04-25  
 ; PRIOR APPLICATION NUMBER: PCT/JP2003/013658  
 ; PRIOR FILING DATE: 2003-10-24  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: PatentIn version 3.3  
 ; SEQ ID NO 2  
 ; LENGTH: 274  
 ; TYPE: PRT  
 ; ORGANISM: Bacillus Licheniformis  
 ; US-10-532-605-2

Query Match 15.5%; Score 532; DB 6; Length 274;

Best Local Similarity 44.6%; Pred. No. 6.6e-30; Indels 36; Gaps 9;

Matches 131; Conservative 37; Mismatches 90; Indels 36; Gaps 9;

QY 145 VMAATVYVNLGDSGITTIGITDIDASHPDLOGKVIQVWVDFVNGRSYDYDHGHTHA 204  
 Db 11 IKADKVOAGQYKAGNVKGIIDTGIASHTDL--KTVGASFYSGSNTYTDGNGHTHV- 67  
 QY 205 STAACTGAASNGK--YKGNAPGAKIAGIVLGDGSGSISTTIIKYEWAVDNKKYKIKY 262  
 Db 68 ---AGTVALDNTTGTGLVAPNVSLYAIKVLNSSGSGTSAISGEMATON---GLDY 120  
 QY 263 INLISGSSSPTDLSQAVNNAMDAQIVVCVAAANGSP--NTYVGSPPAASKIITYG 320  
 Db 121 INNSLID--GPSGSLKQAVDAVASGLVVAANAASSGSGQNTIGYPAKYDSYIANG 177  
 QY 321 AVDSNDINIASFSRGTADGRLEPEVAVAGVDIIAPRAGTSMGTINDYTTKASGTSMA 380  
 Db 178 AVDSNNGRASFSVGS-----ELVMAAGVSVSYSTPNT-----YTSINQTSMA 222  
 QY 381 TTHVSGVALLILOAHPSWTPDKYKIALIETADIAPKEIADIAYGGRVNVYA 434  
 Db 223 SPHVAGAAALILSKYPTLSASQVRNRLSTATNLGD---SFYGGGLINVEAA 272

RESULT 8  
 US-11-452-695-2  
 ; Sequence 2, Application US/11452695  
 ; Publication No. US20060222641A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Neil H. Riordan  
 ; TITLE OF INVENTION: METHOD AND COMPOSITION FOR PREVENTING OR  
 ; FILE REFERENCE: AIDAN.005DV1  
 ; CURRENT APPLICATION NUMBER: US/11/452,695  
 ; CURRENT FILING DATE: 2006-06-14  
 ; PRIOR APPLICATION NUMBER: 60/468948  
 ; PRIOR FILING DATE: 2003-05-07  
 ; PRIOR APPLICATION NUMBER: 10/647131  
 ; PRIOR FILING DATE: 2003-08-22  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 381  
 ; TYPE: PRT  
 ; ORGANISM: Bacillus Subtilis  
 ; US-11-452-695-2

Query Match 15.4%; Score 527; DB 7; Length 381;  
 Best Local Similarity 37.9%; Pred. No. 2.4e-29;  
 Matches 145; Conservative 54; Mismatches 126; Indels 58; Gaps 15;

QY 61 TVIVFBNREKEIAVRVLELMGAKRYRYHHIIPALAADL---KRDVLVISGLTGKAKL 117  
 Db 46 TWSAMSARKKD---VISEKGGKQKQKYNVAAAATLDERAVKEL-----KXD 91  
 QY 118 SGVRFIOEDYKVTSAE-LEGDESAQVMAATYVNNLGYDGSGITIGITDIDASHPDL 176  
 Db 92 PEVAYVEEDH---IAHEVAQSVPYGISQIKAPALHSQGTGSNVKAVVADSGIDSHPD 148  
 QY 177 QCKVIGWDFVNGRSYDYD-HGHGTHVASIAGTGAASNGK--YKGNAPGAKIAGIKYL 233  
 Db 149 --NVAGASFPVSENPYDQDSSHGTHV---AGTIALNNSIGVLGVAAPSILVAVKYL 202  
 QY 234 GADGSGSISTTIIKYEWAVDNKKYKIKYINISLSSGSGSDGTDLSQAVNNAMDAQIV 293  
 Db 203 DSTGSGQYSWIINGIEMALSN---NMDVINNSLG---GPTGSLTKTVVDKAVSSGIV 255  
 QY 294 CVAANGSGP--NTYVGSPPAASKIITYVGAVDNSNDINIASFSRGTADGRLEPEVAPCV 351  
 Db 256 AAAANGSGSGSTVGYPAKPTTAVGAVNSNORASFSSVSGELD-----VNAPEG 309  
 QY 352 DIAPRASGTSMTGPTINDYTTKASGTSMATPHVSGVALLILOAHPSWTPDKYKIALIETA 411  
 Db 310 SIQSTLPQGT-----YGAVNGISMATPHVAGAAALILSKHPTWTVAQVRDRLESTA 360  
 QY 412 DIVAPEIADIAYGGRVNVYA 434  
 Db 361 TYLG---SSFYGGGLINVOAA 379

RESULT 9  
 US-10-541-737-3  
 ; Sequence 3, Application US/10541737  
 ; Publication No. US20060252155A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Leeflang, Chris  
 ; APPLICANT: Van der Kleij, Wilhelmus A.H.  
 ; TITLE OF INVENTION: Methods for Site-Directed Mutagenesis  
 ; FILE REFERENCE: GC787-2  
 ; CURRENT APPLICATION NUMBER: US/10/541,737  
 ; CURRENT FILING DATE: 2005-07-08  
 ; PRIOR APPLICATION NUMBER: PCT/US04/01334  
 ; PRIOR FILING DATE: 2004-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/440,792  
 ; PRIOR FILING DATE: 2003-01-16  
 ; NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 3  
 LENGTH: 274  
 TYPE: PR1  
 ORGANISM: Bacillus licheniformis  
 US-10-541-737-3

Query Match 15.3%, Score 525; DB 6; Length 274;  
 Best Local Similarity 43.5%, Pred. No. 2,1e-29;  
 Matches 126; Conservative 40; Mismatches 90; Indels 36; Gaps 10;

QY 145 VMATYVWNLGYDGGITITGIDTIGIDASHPLDQKVIQWDFVNGRSYPYDGHGTHVA 204  
 DB 11 IKADKVAQGFKGVANVAVLDGTIOASHPDL--NVVGASFFVAGFAVNTDGNHGHV- 67  
 QY 205 SIAGTGAASNGK--YKMAPGAKLAGIKVLGADGSGSISTIIKGVMAVDNKKYKIXV 262  
 DB 68 ---AGTYAALDNTTGVGVAPSVSLVAVKVLNLSGSGSYGIVSGIEMATTN---GMDV 120  
 QY 263 INTSLGSSQSDGTDLSQAVNNMADGIWCVVAGNSG--PNTYVGSPPAASKVITVG 320  
 DB 121 INMSLG---GASGSTAKQAVDANVARGVVAAGNSGSGTNTIGYPAKVDVIAVG 177  
 QY 321 AVDSNDNIASFSSRGPTADGRLKREVVAQGVDTIAPRAGTSMGTPIINDYTTAAGTSM 380  
 DB 178 AVDSNSNRASFSSVG-----ABLEVAPG-----AGVYSTYPTNTYAT--LNGTSM 222  
 QY 381 TPRHVGVALIILQHPMTDPKVKYKALIERADIYAPKEIDIAVAGRVNYYKA 434  
 DB 223 SPHVAGAAAILLSGHPVLSASQVRNRLSTATYIG---SFFYKGLINVEAA 272

RESULT 10  
 US-11-452-695-3

Sequence 3, Application US/11452695  
 Publication No. US20060222641A1  
 GENERAL INFORMATION:  
 APPLICANT: Ne11 H. Riordan  
 TITLE OF INVENTION: METHOD AND COMPOSITION FOR PREVENTING OR  
 FILE REFERENCE: REDUCING DEEP VEIN THROMBOSIS AND/OR PULMONARY EMBOLISM  
 CURRENT APPLICATION NUMBER: US/11/452,695  
 PRIOR FILING DATE: 2006-06-14  
 PRIOR APPLICATION NUMBER: 60/468948  
 PRIOR FILING DATE: 2003-05-07  
 PRIOR APPLICATION NUMBER: 10/647131  
 PRIOR FILING DATE: 2003-08-22  
 NUMBER OF SEQ ID NOS: 7  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 3  
 LENGTH: 381  
 TYPE: PR1  
 ORGANISM: Bacillus Subtilis  
 US-11-452-695-3

Query Match 15.3%, Score 524; DB 7; Length 381;  
 Best Local Similarity 37.9%, Pred. No. 3.9e-29;  
 Matches 145; Conservative 53; Mismatches 127; Indels 58; Gaps 15;

QY 61 TVIVFNHREKEIAVRLVLELMGAKVRYVYHIIIPAIADL---KYRDLVLSGLTGKAKL 117  
 DB 46 TWSAASAKKD---VISEKGVQKQKRYVNAATAITLDEKAVKEL-----KKD 91  
 QY 118 SGVRFIOEDYKVTVAE-LEGIDESAAQVMAVYVNNLGYDGGITITGIDTIGIDASHPL 176  
 DB 92 PSVAAYVEEDH---IAHEVAQSVPGYISQIKAPALHSOGVTGSNVKAVAVDSIGDSHPDL 148  
 QY 177 QGVNIGWDFVNGRSYPYD--HGHTHVASIAAGTGAASNGK--YKMAPGAKLAGIRYL 233  
 DB 149 --NVKGAASFVPSEINPYODGSSHGTHV---AGTIALNNSIGVGVAPASLSLVAVKVL 202  
 QY 234 GADGSGSISTIIKGVMAVDNKKYKIVINLSIGSSQSDGTDLSQAVNNMADGIW 293

DB 203 DSTGSGQSWIINGIEMALISN---NMDVYNLSG---GPTGSLTKTVVDAVSSGIWV 255  
 QY 294 CVAAGNSGP--NTYVGSPPAASKVITVGAVDSDNDNIASFSSRGPTADGRLKREVVAQV 351  
 DB 256 AAAAGNEGSGSISTVGYPAKVPSTIAVGAVNSNORASFSSVGSILD-----VNAQV 309  
 QY 352 DIAPRAGTSMGTPIINDYTTKASGTSMATPHVSGVALIILQHPMTDPKVKYKALIERA 411  
 DB 310 SIOSTLPGT-----YGAVNGTSMATPHVAGAAAILLSKHPVTWNAQVRDLRESTA 360  
 QY 412 DIVAPKEIADIYAGRVNYYKA 434  
 DB 361 TYLG---NSFYKGLINVOAA 379

RESULT 11

US-11-452-695-4  
 Sequence 4, Application US/11452695  
 Publication No. US20060222641A1  
 GENERAL INFORMATION:  
 APPLICANT: Ne11 H. Riordan  
 TITLE OF INVENTION: METHOD AND COMPOSITION FOR PREVENTING OR  
 FILE REFERENCE: REDUCING DEEP VEIN THROMBOSIS AND/OR PULMONARY EMBOLISM  
 CURRENT APPLICATION NUMBER: US/11/452,695  
 PRIOR FILING DATE: 2006-06-14  
 PRIOR APPLICATION NUMBER: 60/468948  
 PRIOR FILING DATE: 2003-05-07  
 PRIOR APPLICATION NUMBER: 10/647131  
 PRIOR FILING DATE: 2003-08-22  
 NUMBER OF SEQ ID NOS: 7  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 4  
 LENGTH: 381  
 TYPE: PR1  
 ORGANISM: Bacillus Subtilis  
 US-11-452-695-4

Query Match 15.3%, Score 524; DB 7; Length 381;  
 Best Local Similarity 37.9%, Pred. No. 3.9e-29;  
 Matches 145; Conservative 53; Mismatches 127; Indels 58; Gaps 15;

QY 61 TVIVFNHREKEIAVRLVLELMGAKVRYVYHIIIPAIADL---KYRDLVLSGLTGKAKL 117  
 DB 46 TWSAASAKKD---VISEKGVQKQKRYVNAATAITLDEKAVKEL-----KKD 91  
 QY 118 SGVRFIOEDYKVTVAE-LEGIDESAAQVMAVYVNNLGYDGGITITGIDTIGIDASHPL 176  
 DB 92 PSVAAYVEEDH---IAHEVAQSVPGYISQIKAPALHSOGVTGSNVKAVAVDSIGDSHPDL 148  
 QY 177 QGVNIGWDFVNGRSYPYD--HGHTHVASIAAGTGAASNGK--YKMAPGAKLAGIRYL 233  
 DB 149 --NVKGAASFVPSEINPYODGSSHGTHV---AGTIALNNSIGVGVAPASLSLVAVKVL 202  
 QY 234 GADGSGSISTIIKGVMAVDNKKYKIVINLSIGSSQSDGTDLSQAVNNMADGIW 293  
 DB 203 DSTGSGQSWIINGIEMALISN---NMDVYNLSG---GPTGSLTKTVVDAVSSGIWV 255  
 QY 294 CVAAGNSGP--NTYVGSPPAASKVITVGAVDSDNDNIASFSSRGPTADGRLKREVVAQV 351  
 DB 256 AAAAGNEGSGSISTVGYPAKVPSTIAVGAVNSNORASFSSVGSILD-----VMAQV 309  
 QY 352 DIAPRAGTSMGTPIINDYTTKASGTSMATPHVSGVALIILQHPMTDPKVKYKALIERA 411  
 DB 310 SIOSTLPGT-----YGAVNGTSMATPHVAGAAAILLSKHPVTWNAQVRDLRESTA 360  
 QY 412 DIVAPKEIADIYAGRVNYYKA 434  
 DB 361 TYLG---NSFYKGLINVOAA 379

RESULT 12

US-11-433-614-16

```
/ Sequence 16, Application US/11433614
/ Publication No. US20060205019A1
/ GENERAL INFORMATION:
/ APPLICANT: Athena Biotechnologies, Inc.
/ TITLE OF INVENTION: Method and Compositions for Directed Gene Assembly
/ FILE REFERENCE: 46675-5004-01-US
/ CURRENT APPLICATION NUMBER: US/11/433,614
/ CURRENT FILING DATE: 2006-05-12
/ PRIOR APPLICATION NUMBER: US 09/920,118
/ PRIOR FILING DATE: 2001-07-31
/ NUMBER OF SEQ ID NOS: 22
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 16
/ LENGTH: 381
/ TYPE: PRN
/ ORGANISM: Bacillus subtilis
US-11-433-614-16

Query Match
Best Local Similarity 37.6%; Pred. No. 1e-28;
Matches 144; Conservative 53; Mismatches 128; Indels 58; Gaps 15;

15.1%; Score 518; DB 7; Length 381;
Query Match
Best Local Similarity 37.6%; Pred. No. 1e-28;
Matches 144; Conservative 53; Mismatches 128; Indels 58; Gaps 15;

QY 61 TVIVENHREKEIAVLELMGAKVRYVYHIIIPALADI--KVRDLVIGLGGKAYL 117
DB 46 TWSAMSAAKKD---VISEKGKVKQKPKYVAAATLDEKAVKEL-----KKD 91

QY 118 SGVRPIQEDYKTVSAE-LEGDESAAOVMATYVNVLYGVDGGITIGITIDGIDASHPTL 176
DB 92 PSVAAYVEEDH---IAHEVYASVPEYGISQIKAPALHSQGYTSNVKAVAVIDSGIDSHPTL 148

QY 177 QGKVIQWVFNVRGSRYPYD-HGHGTHVASIAAGTAANGK--YKGMAPGAKLAGIKYL 233
DB 149 --NRKGASFVSEINPYPYDGGSHGTHV---AGTIALNNSIGVLGAPASLIVAVKYL 202

QY 234 GADGSGISITIIKGVEMAVDNKDKXIKYINLSLSSGSSDGTDSLISOAVNNAMPAGIYV 293
DB 203 DSTGSGQYWIINGIEWAIISN---NMDVYINMSLG---GPTGSLTKTVVDRAVSSGIYV 255

QY 294 CVAANGSG--NTYTGSPAAASKVTYGAVDNSNDINAFSSRGRPTADRLKPEVAVAPV 351
DB 256 AAAAGSGSGSTVGYPAKYPSTIYAVGAVNNSNORASFSSGSEL-----VMAPGV 309

QY 352 DIIAPASGSMGTPIINDYTKASGSMATPHVSGGAILLQAHPSWTPDKYKTLALIERA 411
DB 310 SIQSTLPGST-----YGAINGSMATPHVAAAAALLSKHPYTNAAQVDRLESTRA 360

QY 412 DIVAPKEIADIYAGGRVNVYKA 434
DB 361 TYLG-----NSFYGKGLINVOQA 379

RESULT 13
US-10-567-073-3
/ Sequence 3, Application US/10567073
/ Publication No. US20060134740A1
/ GENERAL INFORMATION:
/ APPLICANT: Bryan, Philip N.
/ TITLE OF INVENTION: Engineered Proteases for Affinity Purification and Processing of
/ FILE REFERENCE: 4115-181
/ CURRENT APPLICATION NUMBER: US/10/567,073
/ CURRENT FILING DATE: 2006-02-03
/ PRIOR APPLICATION NUMBER: US 60/493,032
/ PRIOR FILING DATE: 2003-08-06
/ PRIOR APPLICATION NUMBER: PCT/US04/021049
/ PRIOR FILING DATE: 2004-06-29
/ NUMBER OF SEQ ID NOS: 3
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 3
/ LENGTH: 275
/ TYPE: PRN
/ ORGANISM: Bacillus amyloliquefaciens
```

```
US-10-567-073-3

Query Match
Best Local Similarity 43.3%; Pred. No. 7e-29;
Matches 129; Conservative 41; Mismatches 89; Indels 39; Gaps 11;

15.1%; Score 517.5; DB 6; Length 275;
Query Match
Best Local Similarity 43.3%; Pred. No. 7e-29;
Matches 129; Conservative 41; Mismatches 89; Indels 39; Gaps 11;

QY 143 AQWATYVWNLGYDGGITIGITIDGIDASHPTLQKVIYIWMDFVNGRSYPY-DHGHGT 201
DB 9 SQIKAPALHSQGYTSNVKAVAVIDSGIDSHPTL--KVAAGASWVPSFTNPQDNNSHGT 66

QY 202 HVAASIAAGTAANGK--YKGMAPGAKLAGIKVLGADSGSISTIIKGVEMAVDNKDKXG 259
DB 67 HV-----AGTVAALNNSIGVLGAPASLIVAVKVLGADSGSGYWIINGIEWAIISN-----N 118

QY 202 HVAASIAAGTAANGK--YKGMAPGAKLAGIKVLGADSGSISTIIKGVEMAVDNKDKXG 259
DB 67 HV-----AGTVAALNNSIGVLGAPASLIVAVKVLGADSGSGYWIINGIEWAIISN-----N 118

QY 260 IKVINLSLSSGSSDGTDSLISOAVNNAMPAGIYVCAANGSGP--NTYTGSPAAASKYI 317
DB 119 MDVINMSLG---GPGSAPALKAADVKAASGVVVAALANGESTSGSSTVGPGRKPPSYI 175

QY 318 TVGAVDNSNDINAFSSRGRPTADGRLEPEVAVPGVDIIAPRASGSMGTPIINDYTKASGT 377
DB 176 AVGAVDSSNQRASFSSVGEELD-----VMAPGVSI-----QSTLPGRKYGAYNGT 220

QY 378 SMATPHVSGVGAALLIQAHPSWTPDKYKTLALJETADIAPKEIAD-IAYAGGRVNVYKA 434
DB 221 SVAASPHVGAALAILLSKHPNWTNTQVRSSLENTT-----TKLGSFYYGKGLINVOQA 273

RESULT 14
US-10-541-737-1
/ Sequence 1, Application US/10541737
/ Publication No. US20060252155A1
/ GENERAL INFORMATION:
/ APPLICANT: LeeJiang, Chris
/ TITLE OF INVENTION: Methods for Site-Directed Mutagenesis
/ FILE REFERENCE: GC787-2
/ CURRENT APPLICATION NUMBER: US/10/541,737
/ CURRENT FILING DATE: 2005-07-08
/ PRIOR APPLICATION NUMBER: PCT/US04/01334
/ PRIOR FILING DATE: 2004-01-16
/ PRIOR APPLICATION NUMBER: US 60/440,792
/ PRIOR FILING DATE: 2003-01-16
/ NUMBER OF SEQ ID NOS: 6
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 275
/ TYPE: PRN
/ ORGANISM: Bacillus amyloliquefaciens
US-10-541-737-1

Query Match
Best Local Similarity 15.1%; Score 517.5; DB 6; Length 275;
Matches 129; Conservative 41; Mismatches 89; Indels 39; Gaps 11;

15.1%; Score 517.5; DB 6; Length 275;
Query Match
Best Local Similarity 43.3%; Pred. No. 7e-29;
Matches 129; Conservative 41; Mismatches 89; Indels 39; Gaps 11;

QY 143 AQWATYVWNLGYDGGITIGITIDGIDASHPTLQKVIYIWMDFVNGRSYPY-DHGHGT 201
DB 9 SQIKAPALHSQGYTSNVKAVAVIDSGIDSHPTL--KVAAGASWVPSFTNPQDNNSHGT 66

QY 202 HVAASIAAGTAANGK--YKGMAPGAKLAGIKVLGADSGSISTIIKGVEMAVDNKDKXG 259
DB 67 HV-----AGTVAALNNSIGVLGAPASLIVAVKVLGADSGSGYWIINGIEWAIISN-----N 118

QY 260 IKVINLSLSSGSSDGTDSLISOAVNNAMPAGIYVCAANGSGP--NTYTGSPAAASKYI 317
DB 119 MDVINMSLG---GPGSAPALKAADVKAASGVVVAALANGESTSGSSTVGPGRKPPSYI 175

QY 318 TVGAVDNSNDINAFSSRGRPTADGRLEPEVAVPGVDIIAPRASGSMGTPIINDYTKASGT 377
DB 176 AVGAVDSSNQRASFSSVGEELD-----VMAPGVSI-----QSTLPGRKYGAYNGT 220

QY 378 SMATPHVSGVGAALLIQAHPSWTPDKYKTLALJETADIAPKEIAD-IAYAGGRVNVYKA 434
DB 221 SVAASPHVGAALAILLSKHPNWTNTQVRSSLENTT-----TKLGSFYYGKGLINVOQA 273
```

DB 221 SMASPHVAGAAALILSKHPMTNTQVRSLENTT-----TKLGDSFYKGLINVQAA 273

RESULT 15

US-10-526-324-20

Sequence 20, Application US/10526324

Publication No. US20060248617A1

GENERAL INFORMATION:

APPLICANT: Imanaka, Takayuki

APPLICANT: Atomi, Haruyuki

TITLE OF INVENTION: METHOD OF TARGETED GENE DISRUPTION, GENOME OF

TITLE OF INVENTION: HYPERTHERMOSTABLE BACTERIUM AND GENOME CHIP USING

TITLE OF INVENTION: THE SAME

FILE REFERENCE: 490051.401USPC

CURRENT APPLICATION NUMBER: US/10/526,324

PRIOR APPLICATION NUMBER: PCT/IB2003/003597

PRIOR FILING DATE: 2003-08-29

PRIOR APPLICATION NUMBER: JP 2002-319011

PRIOR FILING DATE: 2002-08-30

NUMBER OF SEQ ID NOS: 2167

SOFTWARE: PatentIn version 3.1

SEQ ID NO 20

LENGTH: 524

TYPE: PRT

ORGANISM: Thermococcus kodakarensis KOD1

FEATURE:

NAME/KEY: misc feature

LOCATION: (414542)..(414542)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (786890)..(786890)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (786907)..(786907)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (786944)..(786944)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (839139)..(839139)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1128488)..(1128488)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1128499)..(1128499)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1128505)..(1128505)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1128517)..(1128517)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1128539)..(1128539)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1561400)..(1561400)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1561477)..(1561477)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1561545)..(1561545)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1767941)..(1767941)

OTHER INFORMATION: n is a or c or g or t.

US-10-526-324-20

Query Match 14.7%; Score 504.5; DB 6; Length 524;

Best Local Similarity 35.9%; Pred. No. 1.5e-27;

Matches 157; Conservative 67; Mismatch 154; Indels 59; Gaps 18;

QY 51 OKLNPERISTIVF---ENHREKEIAVAVLEIMGAKVYVYHIIIPAIADLKVRLVY 107

DB 104 QKVONDEPIQWIFITTKPNDRKE---ALEHIGAEVLYVDDDAIIS--LKARISL-I 156

QY 108 SGLTGKAKLGGVRF-IQEDY-----KTVSAR-LRGLDESAQ 144

DB 157 KNLVVSQAFDPDYRFIRVWMPDLVETGPTILENETFTTPANVITTPROVSGAINIKL 216

QY 145 VVATYVWNL-GYDGSITIGIIDTGIDASHPDLOGKVIGWVDFVNGRSYFYDD-HGHGTH 202

DB 217 VKADILMSKRGITGKGVVAVLDTGVDCHVMLOGACVFENFVDE--PAKDNLNGGTH 274

QY 203 VASIAAG--TGAASNGK---YKGMAPGAKLAGIKVLGADSGSISTIIKGVENAVDNKDK 257

DB 275 VAGIINAGPTKYTWBGEKVVYVSGVAPEANILAVKVGQGGGTMTIIGLDIVVWKKK 334

QY 258 YGIR--VINLSLSSQSOSDGTDSLQAANNMADGIVCVAAGNSGPNYTVGSPAAAK 315

DB 335 HPEEPVIVMSLGSPPGSPRDPVQVBOIIREHILPVYIAGN---BFAVIDSPGIATG 391

QY 316 VITVGAVDNDNIAGFSRGPFTAD-GRLEPVVAPGVDTIAPASGTSMTPIINDYTTA 374

DB 392 AITVAAVDNNMKVASFSGKPGIINIVDIRPIDIAAPVKILSARA---GT--RNEFIAM 444

QY 375 SGTSMATPHVSGVAGALILQAHPSWTPDKVTKLILIEPTADIVAPRSIADIVAGRVVYTKA 434

DB 445 SGTSMATPHVSGVAGALILQAHGDLTPETIKLILEKTAIYPLDGLDLPVSGAGVDAIYAA 504

QY 435 IKYDDYAKLTFGTSVAD 451

DB 505 VK---ABPSBSGGLTD 517

Search completed: January 6, 2007, 22:38:59

Job time : 33 secs



GenCore version 5.1.9  
Copyright (c) 1993 - 2007 Bioacceleration Ltd.

OW protein - protein search, using sw model

Run on: January 6, 2007, 22:31:42 ; Search time 36 Seconds

(Without alignments)  
1602.297 Million cell updates/sec

Title: US-10-800-684-1

Perfect score: 3437

Sequence: 1 MKRLGAVLVVGLACT.....YAVSYGMADYQKAVVYVG 659

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /EMC\_Celerra\_SIDS3/prodata/2/1aa/5/COMB.pep:\*  
2: /EMC\_Celerra\_SIDS3/prodata/2/1aa/6/COMB.pep:\*  
3: /EMC\_Celerra\_SIDS3/prodata/2/1aa/7/COMB.pep:\*  
4: /EMC\_Celerra\_SIDS3/prodata/2/1aa/H/COMB.pep:\*  
5: /EMC\_Celerra\_SIDS3/prodata/2/1aa/PCTUS.COMB.pep:\*  
6: /EMC\_Celerra\_SIDS3/prodata/2/1aa/RE.COMB.pep:\*  
7: /EMC\_Celerra\_SIDS3/prodata/2/1aa/backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description       |
|------------|--------|-------------|--------|-------|-------------------|
| 1          | 3437   | 100.0       | 659    | 2     | US-08-894-818B-1  |
| 2          | 3437   | 100.0       | 659    | 2     | US-09-445-472-12  |
| 3          | 3437   | 100.0       | 659    | 2     | US-10-090-624-12  |
| 4          | 3437   | 100.0       | 659    | 2     | US-09-841-553-1   |
| 5          | 2914   | 84.8        | 659    | 2     | US-08-894-818B-5  |
| 6          | 2914   | 84.8        | 659    | 2     | US-09-841-553-5   |
| 7          | 2483.5 | 72.3        | 654    | 2     | US-08-894-818B-35 |
| 8          | 2483.5 | 72.3        | 654    | 2     | US-09-445-472-16  |
| 9          | 2483.5 | 72.3        | 654    | 2     | US-10-090-624-16  |
| 10         | 2483.5 | 72.3        | 654    | 2     | US-09-841-553-35  |
| 11         | 2138.5 | 62.2        | 522    | 2     | US-08-894-818B-3  |
| 12         | 2138.5 | 62.2        | 522    | 2     | US-09-445-472-4   |
| 13         | 2138.5 | 62.2        | 522    | 2     | US-10-090-624-4   |
| 14         | 2138.5 | 62.2        | 522    | 2     | US-09-841-553-3   |
| 15         | 1707   | 49.7        | 412    | 2     | US-09-445-472-1   |
| 16         | 1707   | 49.7        | 412    | 2     | US-10-090-624-1   |
| 17         | 1148   | 33.4        | 237    | 1     | US-08-750-532-18  |
| 18         | 665    | 19.3        | 734    | 2     | US-09-000-016-4   |
| 19         | 665    | 19.3        | 734    | 2     | US-09-514-340-4   |
| 20         | 665    | 19.3        | 823    | 2     | US-09-000-016-2   |
| 21         | 665    | 19.3        | 823    | 2     | US-09-514-340-2   |
| 22         | 647.5  | 18.8        | 520    | 2     | US-09-000-016-7   |
| 23         | 647.5  | 18.8        | 520    | 2     | US-09-514-340-7   |
| 24         | 560.5  | 16.2        | 379    | 1     | US-08-685-774-2   |
| 25         | 557.5  | 16.2        | 350    | 1     | US-07-923-260A-4  |
| 26         | 557.5  | 16.2        | 379    | 1     | US-08-845-295A-1  |

|    |       |      |      |   |                  |                    |
|----|-------|------|------|---|------------------|--------------------|
| 27 | 557.5 | 16.2 | 379  | 2 | US-09-140-933-1  | Sequence 1, Appl1  |
| 28 | 557.5 | 16.2 | 379  | 2 | US-09-146-661-1  | Sequence 1, Appl1  |
| 29 | 557.5 | 16.2 | 379  | 2 | US-09-150-515-1  | Sequence 1, Appl1  |
| 30 | 553   | 16.1 | 378  | 1 | US-07-772-087-4  | Sequence 4, Appl1  |
| 31 | 535.5 | 15.6 | 1398 | 1 | US-08-750-532-9  | Sequence 9, Appl1  |
| 32 | 535.5 | 15.6 | 1398 | 2 | US-08-894-818B-8 | Sequence 8, Appl1  |
| 33 | 535.5 | 15.6 | 1398 | 2 | US-09-445-472-6  | Sequence 6, Appl1  |
| 34 | 535.5 | 15.6 | 1398 | 2 | US-10-090-624-6  | Sequence 6, Appl1  |
| 35 | 535.5 | 15.5 | 1398 | 2 | US-09-841-553-8  | Sequence 8, Appl1  |
| 36 | 533.5 | 15.5 | 352  | 1 | US-07-923-260A-2 | Sequence 2, Appl1  |
| 37 | 533.5 | 15.5 | 382  | 2 | US-09-255-502-2  | Sequence 2, Appl1  |
| 38 | 533.5 | 15.5 | 382  | 2 | US-09-445-472-31 | Sequence 31, Appl1 |
| 39 | 533.5 | 15.5 | 382  | 2 | US-09-659-749-2  | Sequence 2, Appl1  |
| 40 | 533.5 | 15.5 | 382  | 2 | US-09-672-105-2  | Sequence 2, Appl1  |
| 41 | 533.5 | 15.5 | 382  | 2 | US-10-090-624-31 | Sequence 31, Appl1 |
| 42 | 533.5 | 15.5 | 382  | 2 | US-10-104-693-2  | Sequence 2, Appl1  |
| 43 | 532.5 | 15.5 | 382  | 2 | US-09-178-155-2  | Sequence 2, Appl1  |
| 44 | 532.5 | 15.5 | 382  | 2 | US-09-178-173A-2 | Sequence 2, Appl1  |
| 45 | 532.5 | 15.5 | 382  | 2 | US-09-177-353-2  | Sequence 2, Appl1  |

#### ALIGNMENTS

RESULT 1  
US-08-894-818B-1  
Sequence 1, Application US/08894818B  
Patent No. 6261822  
GENERAL INFORMATION:  
APPLICANT: TAKAKURA, Hikaru  
APPLICANT: MORISHITA, Mio  
APPLICANT: YAMAMOTO, Katsuniko  
APPLICANT: MITTA, Masanori  
APPLICANT: ASADA, Kiyozo  
APPLICANT: TSUNASAWA, Susumu  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street N.W., Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/894, 818B  
FILING DATE: 20-MAY-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP96/03253  
FILING DATE: 07-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 323285/1995  
FILING DATE: 12-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TAKAKURA=1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ. ID NO. 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 659 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

APP-THIS COPY  
ANV

MOLECULE TYPE: peptide  
US-08-894-8188-1

Query Match 100.0%; Score 3437; DB 2; Length 659;  
Best Local Similarity 100.0%; Pred. No. 1.9e-236;  
Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MKRLGAVVLAIVLVLGAGTALAAAPVPRVNNNAVQOKNYGLTLPGLFKYQORNNNOEV 60
DB 1 MKRLGAVVLAIVLVLGAGTALAAAPVPRVNNNAVQOKNYGLTLPGLFKYQORNNNOEV 60
QY 61 DTVMFGSYGDRRAVAVLRLMGAQVYKIIIPAAVKIKARDLLIAGMIDTGYFGNT 120
DB 61 DTVMFGSYGDRRAVAVLRLMGAQVYKIIIPAAVKIKARDLLIAGMIDTGYFGNT 120
QY 121 RVSGIKFIQEDYKQVVDATSVSQIGADTVNNSLGYDGSVVVAIVDTGIDANHPDLKKG 180
DB 121 RVSGIKFIQEDYKQVVDATSVSQIGADTVNNSLGYDGSVVVAIVDTGIDANHPDLKKG 180
QY 181 VIGMYDAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGS 240
DB 181 VIGMYDAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGS 240
QY 241 VSTIIAGVDMVQNKDKYGRVNLISLGSQSSDGTDSLQAVNNAMDAGIIVCVAAGNS 300
DB 241 VSTIIAGVDMVQNKDKYGRVNLISLGSQSSDGTDSLQAVNNAMDAGIIVCVAAGNS 300
QY 301 GPNTYVGSPPAAASKVITVGAVDNDNIASFSSRGPTADGRLEKVVAPGVDIIPRAGS 360
DB 301 GPNTYVGSPPAAASKVITVGAVDNDNIASFSSRGPTADGRLEKVVAPGVDIIPRAGS 360
QY 361 TSMGTPINDYTTKASGTSMTAPHSVGVALILQAHSPMTDPKVTALIFADIAPKEIA 420
DB 361 TSMGTPINDYTTKASGTSMTAPHSVGVALILQAHSPMTDPKVTALIFADIAPKEIA 420
QY 421 DIAYGARNVYKAIKYDYAKLFTGSAVDKGSATHTFDVSGATFTVATLTYDTSDDI 480
DB 421 DIAYGARNVYKAIKYDYAKLFTGSAVDKGSATHTFDVSGATFTVATLTYDTSDDI 480
QY 481 DLVLYDPNGNEVDYSTAYYGFKEVGYNPTAGTWTVKVVSYKGAANYQVDVSDGSLQ 540
DB 481 DLVLYDPNGNEVDYSTAYYGFKEVGYNPTAGTWTVKVVSYKGAANYQVDVSDGSLQ 540
QY 541 SGGNNPNPNPNPTPTTDTQTFTGSVNDYWDTSDFTMNVNSGATKLTGDLTFDTSYND 600
DB 541 SGGNNPNPNPNPTPTTDTQTFTGSVNDYWDTSDFTMNVNSGATKLTGDLTFDTSYND 600
QY 601 LDLYLDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVYAYSTYGMADYQLKAVVYYG 659
DB 601 LDLYLDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVYAYSTYGMADYQLKAVVYYG 659
```

## RESULT 2

US-09-445-472-12  
Sequence 12, Application US/09445472  
Patent No. 6358726  
GENERAL INFORMATION:  
APPLICANT: TAKAKURA, Hikaru  
APPLICANT: MORISHITA, Mio  
APPLICANT: SHIMOTO, Tomoko  
APPLICANT: ASADA, Kiyozo  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
FILE REFERENCE: TAKAKURA=6  
CURRENT APPLICATION NUMBER: US/09/445,472  
CURRENT FILING DATE: 1999-12-06  
PRIOR APPLICATION NUMBER: 151969/1997  
PRIOR FILING DATE: 1997-06-10  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 12  
LENGTH: 659  
TYPE: PRT

No bbl part  
No bbl part

ORGANISM: Thermococcus celer  
US-09-445-472-12

Query Match 100.0%; Score 3437; DB 2; Length 659;  
Best Local Similarity 100.0%; Pred. No. 1.9e-236;  
Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MKRLGAVVLAIVLVLGAGTALAAAPVPRVNNNAVQOKNYGLTLPGLFKYQORNNNOEV 60
DB 1 MKRLGAVVLAIVLVLGAGTALAAAPVPRVNNNAVQOKNYGLTLPGLFKYQORNNNOEV 60
QY 61 DTVMFGSYGDRRAVAVLRLMGAQVYKIIIPAAVKIKARDLLIAGMIDTGYFGNT 120
DB 61 DTVMFGSYGDRRAVAVLRLMGAQVYKIIIPAAVKIKARDLLIAGMIDTGYFGNT 120
QY 121 RVSGIKFIQEDYKQVVDATSVSQIGADTVNNSLGYDGSVVVAIVDTGIDANHPDLKKG 180
DB 121 RVSGIKFIQEDYKQVVDATSVSQIGADTVNNSLGYDGSVVVAIVDTGIDANHPDLKKG 180
QY 181 VIGMYDAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGS 240
DB 181 VIGMYDAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGS 240
QY 241 VSTIIAGVDMVQNKDKYGRVNLISLGSQSSDGTDSLQAVNNAMDAGIIVCVAAGNS 300
DB 241 VSTIIAGVDMVQNKDKYGRVNLISLGSQSSDGTDSLQAVNNAMDAGIIVCVAAGNS 300
QY 301 GPNTYVGSPPAAASKVITVGAVDNDNIASFSSRGPTADGRLEKVVAPGVDIIPRAGS 360
DB 301 GPNTYVGSPPAAASKVITVGAVDNDNIASFSSRGPTADGRLEKVVAPGVDIIPRAGS 360
QY 361 TSMGTPINDYTTKASGTSMTAPHSVGVALILQAHSPMTDPKVTALIFADIAPKEIA 420
DB 361 TSMGTPINDYTTKASGTSMTAPHSVGVALILQAHSPMTDPKVTALIFADIAPKEIA 420
QY 421 DIAYGARNVYKAIKYDYAKLFTGSAVDKGSATHTFDVSGATFTVATLTYDTSDDI 480
DB 421 DIAYGARNVYKAIKYDYAKLFTGSAVDKGSATHTFDVSGATFTVATLTYDTSDDI 480
QY 481 DLVLYDPNGNEVDYSTAYYGFKEVGYNPTAGTWTVKVVSYKGAANYQVDVSDGSLQ 540
DB 481 DLVLYDPNGNEVDYSTAYYGFKEVGYNPTAGTWTVKVVSYKGAANYQVDVSDGSLQ 540
QY 541 SGGNNPNPNPNPTPTTDTQTFTGSVNDYWDTSDFTMNVNSGATKLTGDLTFDTSYND 600
DB 541 SGGNNPNPNPNPTPTTDTQTFTGSVNDYWDTSDFTMNVNSGATKLTGDLTFDTSYND 600
QY 601 LDLYLDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVYAYSTYGMADYQLKAVVYYG 659
DB 601 LDLYLDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVYAYSTYGMADYQLKAVVYYG 659
```

## RESULT 3

US-10-090-624-12  
Sequence 12, Application US/10090624  
Patent No. 6783970  
GENERAL INFORMATION:  
APPLICANT: TAKAKURA, Hikaru  
APPLICANT: MORISHITA, Mio  
APPLICANT: SHIMOTO, Tomoko  
APPLICANT: ASADA, Kiyozo  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
FILE REFERENCE: TAKAKURA=6  
CURRENT APPLICATION NUMBER: US/10/090,624  
CURRENT FILING DATE: 2002-03-06  
PRIOR APPLICATION NUMBER: 09/445,472  
PRIOR FILING DATE: 1999-12-06  
PRIOR APPLICATION NUMBER: 151969/1997  
PRIOR FILING DATE: 1997-06-10  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 12

No bbl part  
No bbl part

```

; LENGTH: 659
; TYPE: prt
; ORGANISM: Thermococcus celer
US-10-090-624-12

```

|                       |                 |                     |           |             |
|-----------------------|-----------------|---------------------|-----------|-------------|
| Query Match           | 100.0%;         | Score 3437;         | DB 2;     | Length 659; |
| Best Local Similarity | 100.0%;         | Pred. No. 1.9e-236; |           |             |
| Matches 659;          | Conservative 0; | Mismatches 0;       | Indels 0; | Gaps 0;     |

|    |     |   |     |
|----|-----|---|-----|
| Qy | 1   | MKRLGAVLALVVLVGLLAGLALAAPEKPAVNNNAVQAKYGLITGELFKYQRMNNQEV     | 60  |
| Db | 1   | MKRLGAVLALVVLVGLLAGLALAAPEKPAVNNNAVQAKYGLITGELFKYQRMNNQEV     | 60  |
| Qy | 61  | DTYIMFGSTGDRDRAVKRLIMGADQYKYSYKIIPAVAVKIKARDILLIAGMITDGYEGNT  | 120 |
| Db | 61  | DTYIMFGSGYDREDRAVKRLIMGADQYKYSYKIIPAVAVKIKARDILLIAGMITDGYEGNT | 120 |
| Qy | 121 | RVSGIKFIEDKVKVQVDDATSVSOIGADPVNMSLGIDSSGVVVALVPIGDIAHPDLKKG   | 180 |
| Db | 121 | RVSGIKFIEDKVKVQVDDATSVSOIGADPVNMSLGIDSSGVVVALVPIGDIAHPDLKKG   | 180 |
| Qy | 181 | VIGWYDVAVNGRSTPYDDQGHGTHVAGIVAGTSSVNSQYIGVAPGAKLVGVKVLGADSGGS | 240 |
| Db | 181 | VIGWYDVAVNGRSTPYDDQGHGTHVAGIVAGTSSVNSQYIGVAPGAKLVGVKVLGADSGGS | 240 |
| Qy | 241 | VSTIIIGDVVVQONKDKYIGRIVINISLGSSGSGSDGTDLSLQAVNNMADGIIVCAAQNS  | 300 |
| Db | 241 | VSTIIIGDVVVQONKDKYIGRIVINISLGSSGSGSDGTDLSLQAVNNMADGIIVCAAQNS  | 300 |
| Qy | 301 | GPRTYVGSPPAASKYITVGAVDSDNNTASFSSRPTDAGRLKPEVPAAGVDIIAPASG     | 360 |
| Db | 301 | GPRTYVGSPPAASKYITVGAVDSDNNTASFSSRPTDAGRLKPEVPAAGVDIIAPASG     | 360 |
| Qy | 361 | TSMGTEINDYTKASGTSNATPHVSGVALIIQAHPSWTPDKVKVALLIETADIVAPKEIA   | 420 |
| Db | 361 | TSMGTEINDYTKASGTSNATPHVSGVALIIQAHPSWTPDKVKVALLIETADIVAPKEIA   | 420 |
| Qy | 421 | DIAYGGRVNVYKAIKYDYAKLITFGSYADBGSAITHFDVSGATFVTAALYMTGSSDI     | 480 |
| Db | 421 | DIAYGGRVNVYKAIKYDYAKLITFGSYADBGSAITHFDVSGATFVTAALYMTGSSDI     | 480 |
| Qy | 481 | DLVLYPNGNEVDYSTAYYGFEEKGYNPTAGTMTVVVSYKGAANYQVQVWVDSGLSQ      | 540 |
| Db | 481 | DLVLYPNGNEVDYSTAYYGFEEKGYNPTAGTMTVVVSYKGAANYQVQVWVDSGLSQ      | 540 |
| Qy | 541 | SGGGENPNPNPPTPTDQTGTFGGSVNDYMDSDTLTMNNSGATKLTGDLTDTDSYND      | 600 |
| Db | 541 | SGGGENPNPNPPTPTDQTGTFGGSVNDYMDSDTLTMNNSGATKLTGDLTDTDSYND      | 600 |
| Qy | 601 | LDDLYLIDPNNGNLVDRSTSSNSYEHAYEPANPAGTMTFLVYAYSTYGMAVDOLKAAVYVG | 659 |
| Db | 601 | LDDLYLIDPNNGNLVDRSTSSNSYEHAYEPANPAGTMTFLVYAYSTYGMAVDOLKAAVYVG | 659 |

[illegible]

RESULT 4  
 US-09-841-553-1  
 ; Sequence 1, Application US/09841553  
 ; Patent No. 6849441  
 ;  
 GENERAL INFORMATION:  
 APPLICANT: TAKAKURA, Hikaru  
 MORISHITA, Mio  
 YAMAMOTO, Katsuhiko  
 MITTA, Masamori  
 ASADA, Kiyozo  
 TSUNASAWA, Susumu  
 KATO, Ikumoshin  
 TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
 NUMBER OF SEQUENCES: 42  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Broadway and Netmark  
 STREET: 419 Seventh Street N.W., Ste. 300  
 CITY: Washington  
 STATE: D.C.

553-1  
 553-1

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

|                       |                 |  |           |             |
|-----------------------|-----------------|--|-----------|-------------|
| Query Match           | 100.0 %;        | Score 3437;  | DB 2;     | Length 659; |
| Best Local Similarity | 100.0 %;        | Pred. No. 1.9e-236;  |           |             |
| Matches 659;          | Conservative 0; | Mismatches 0;  | Indels 0; | Gaps 0;     |
| QY                    | 1               | MKRLGAVVIALVILVGLIAGTALAAAPVAPVVRNNAVVOQKNTGLLTPGLFKYQRMNNMVEV   | 60        |             |
| Db                    | 1               | MKRLGAVVIALVILVGLIAGTALAAAPVAPVVRNNAVVOQKNTGLLTPGLFKYQRMNNMVEV   | 60        |             |
| QY                    | 61              | DTYIMFSSYEDRDBRAVKVLRIMGQVKKYISKIIPAVAVKIKARDLLLIAGMIDTGYRGNT    | 120       |             |
| Db                    | 61              | DTYIMFSSYEDRDBRAVKVLRIMGQVKKYISKIIPAVAVKIKARDLLLIAGMIDTGYRGNT    | 120       |             |
| QY                    | 121             | RVSGIKRIGEDYKQVDDATSVSOGIDATYWNLSIGYSGSGVVAIVDTGIDANHPDLKX       | 180       |             |
| Db                    | 121             | RVSGIKRIGEDYKQVDDATSVSOGIDATYWNLSIGYSGSGVVAIVDTGIDANHPDLKX       | 180       |             |
| QY                    | 181             | VIGMYDAVNRSTPYDDQGHGTNAGIIVAGTGSVNSQYIGVAPAKLVGVKVLGADSGGS       | 240       |             |
| Db                    | 181             | VIGMYDAVNRSTPYDDQGHGTNAGIIVAGTGSVNSQYIGVAPAKLVGVKVLGADSGGS       | 240       |             |
| QY                    | 241             | VSTIIAGVDVNVONCKDKYGRVITNLSIGSSQSSDGTDSLSDQAVNNAMDAGIIVCCVAAAGNS | 300       |             |
| Db                    | 241             | VSTIIAGVDVNVONCKDKYGRVITNLSIGSSQSSDGTDSLSDQAVNNAMDAGIIVCCVAAAGNS | 300       |             |
| QY                    | 301             | GPMTTYTGSPPAAASKVITTVGAVVNSNDIAFSSSGPAPDRKIREVVAPGVDIIAPRASG     | 360       |             |
| Db                    | 301             | GPMTTYTGSPPAAASKVITTVGAVVNSNDIAFSSSGPAPDRKIREVVAPGVDIIAPRASG     | 360       |             |
| QY                    | 361             | TSMGTPINDYYTASGTSMAATPHVSGGALILQAPHSWTEPDKVKTALILETDIIVAPKEIA    | 420       |             |
| Db                    | 361             | TSMGTPINDYYTASGTSMAATPHVSGGALILQAPHSWTEPDKVKTALILETDIIVAPKEIA    | 420       |             |
| QY                    | 421             | DIYAGAGRVNVYKAIKTDYAKLITFGSVADKSGATHTFDVSGATFTYATLTYMDTSSDI      | 480       |             |
| Db                    | 421             | DIYAGAGRVNVYKAIKTDYAKLITFGSVADKSGATHTFDVSGATFTYATLTYMDTSSDI      | 480       |             |
| QY                    | 481             | DLVLYPBNGENEVYSYATAYYGFEVGVYNNPAGTWTWKVSYKGAANYQVDVSDGSLSQ       | 540       |             |
| Db                    | 481             | DLVLYPBNGENEVYSYATAYYGFEKVGYNPAGTWTWKVSYKGAANYQVDVSDGSLSQ        | 540       |             |
| QY                    | 541             | SGGAGNPENPNPPTTDTQTGTGSVNDYMDTSDTFPMNVNSGATKITGDLTFDTSYND        | 600       |             |

Db 541 SGGGNPNPNPPTPTDPTOTFTGSGVNDWTDSTFTMNVNSGATKITDGLTFDTSYND 600  
QY 601 LDIYLYDPNGNLVDRSTSSNSYEHEVEYANPAPGWTFLVAYASTYGMADYQLAVVYVG 659  
Db 601 LDIYLYDPNGNLVDRSTSSNSYEHEVEYANPAPGWTFLVAYASTYGMADYQLAVVYVG 659

RESULT 5  
US-08-894-818B-5  
Sequence 5, Application US/08894818B  
Patent No. 6261822  
GENERAL INFORMATION:  
APPLICANT: TAKAKURA, Hikaru  
APPLICANT: MORISHITA, Mio  
APPLICANT: YAMAMOTO, Katsuhiko  
APPLICANT: MITTA, Masanori  
APPLICANT: ASADA, Kiyoza  
APPLICANT: TSUNASAWA, Susumu  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street N.W., Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/894,818B  
FILING DATE: 20-MAY-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP96/03253  
FILING DATE: 07-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 323285/1995  
FILING DATE: 12-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TAKAKURA-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SRO ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 659 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-894-818B-5

Query Match 84.8%; Score 2914; DB 2; Length 659;  
Best Local Similarity 83.7%; Pred. No. 3.2e-199;  
Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps 7;

QY 1 MKRLGAVTALVLVGLAGTALAAPVTPV--VRNNAAVOQKQYGLITGLFRKQVQRNMNQ 58  
Db 1 MKGKALILVILVGLVGVGAAPBEKKVQVRN---VERKYGILITGLFRKIKLNPNE 57  
QY 59 EVDVIMFGSGVGDDBRAVKVLRMLGAQVKSYKIIIPAVAKIKARDLLLITAGMDTGYFG 118  
Db 58 EISTVIVFENHREKEIIVALELGAQVKRYVYHIIIPALADKLVLISGL--TS--G 113  
QY 119 NTRVSGIKFTQBDYKQVDDA----TSVSGIGADTVWNSLGYDGSVVVAIVDTGIDAN 173

Db 114 KATLSGVRFTQBDYKQVTSAELEGLDESSAQVATVYWN--LGVDGSGITITGIIDTGDAS 172  
QY 174 HPLDKGVIGWYAVNAGRSTPYDDQGHGVAGIVAGTGSV--NSQYIGVAPGAKLVGVKV 232  
Db 173 HPLDQGVIGWYVFNGRSTPYDDHGHGVASIAAGTGAASNGKIKGAAPGAKLAGIKV 232  
QY 233 LGADGSGSVSTIIAGVWVYVQNKRYGIRVINLSLGSQSGSDGTDLSLQAVNNAMPAGIV 292  
Db 233 LGADGSGSISTIIKGVWAVNDKDKYGIKVINLSLGSQSGSDGTDLSLQAVNNAMPAGIV 292  
QY 293 VCVAAAGNSGNTTYSPPAAASRVITVGAVDSNDNLSFSSRPTADGRUKPEVVAAGVD 352  
Db 293 VCVAAAGNSGNTTYSPPAAASRVITVGAVDSNDNLSFSSRPTADGRUKPEVVAAGVD 352  
QY 353 IIPRASGSMGPIINDYTTKASGTSMATPHVSGVALILQAPSPMPDKVKTALLETAD 412  
Db 353 IIPRASGSMGPIINDYTTKASGTSMATPHVSGVALILQAPSPMPDKVKTALLETAD 412  
QY 413 IAPKEIADIAYGAGRVNYKAIKYDDYAKLFTGSAVDKGSATHTFDVSGATFTATLY 472  
Db 413 IAPKEIADIAYGAGRVNYKAIKYDDYAKLFTGSAVDKGSATHTFDVSGATFTATLY 472  
QY 473 WDTGSSDIDLXYDPNGNEVDYSYTAAYGFEKYGYNPTAGTWTYKVSYKGAANYQVDV 532  
Db 473 WDTGSSDIDLXYDPNGNEVDYSYTAAYGFEKYGYNPTAGTWTYKVSYKGAANYQVDV 532  
QY 533 VSDGSLSGSGGPNPNPNPPTPTDPTOTFTGSGVNDWTDSTFTMNVNSGATKITDGL 592  
Db 533 VSDGSLSGSGGPNPNPNPPTPTDPTOTFTGSGVNDWTDSTFTMNVNSGATKITDGL 592  
QY 593 TPDTSYNDLXYLYDPNGNLVDRSTSSNSYEHEVEYANPAPGWTFLVAYASTYGMADYQL 652  
Db 593 TPDTSYNDLXYLYDPNGNLVDRSTSSNSYEHEVEYANPAPGWTFLVAYASTYGMADYQL 652  
QY 653 KAVVYVG 659  
Db 653 KAVVYVG 659

RESULT 6  
US-09-841-553-5  
Sequence 5, Application US/09841553  
Patent No. 6849441  
GENERAL INFORMATION:  
APPLICANT: TAKAKURA, Hikaru  
APPLICANT: MORISHITA, Mio  
APPLICANT: YAMAMOTO, Katsuhiko  
APPLICANT: MITTA, Masanori  
APPLICANT: ASADA, Kiyoza  
APPLICANT: TSUNASAWA, Susumu  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street N.W., Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/841,553  
FILING DATE: 24-APR-2001  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/894,818  
FILING DATE: <Unknown>  
APPLICATION NUMBER: JP 323285/1995

FILING DATE: 12-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TAKAKURA-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 5  
SEQUENCE CHARACTERISTICS:  
LENGTH: 659 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-841-553-5

Query Match 84.8%; Score 2914; DB 2; Length 659;  
Best Local Similarity 83.7%; Pred. No. 3,2e-199;  
Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps 7;

QY 1 MRRLGAVTLALVGLLALTAAPKPV--VRNNAVQOKNYGLTPGLFKYQRMNMQ 58  
DB 1 MGKLKALILVILVLGLVGSVAAPKRYQVRN--VERKYGILTPGLFRKIQKLNPE 57  
QY 59 EVDVTIMFGSYGDRRAVAVLRMGAVKYSYKIIIPAVNKKIRADLLIAGMIDGYRG 118  
DB 58 EISTVIVFENHREKEIAVRVLELMGAKRVYVHIIPALADLKVRDLVLSGL--TG--G 113  
QY 119 NTRVSGIKFIQEDYKQVDDA-----TSVSIQADTVWNSLGYDSGVAIVDTGIDAN 173  
DB 114 KAKLSGVRFIQEYKXTVSALEGLDESAQVWATYVN--LGYDSGIGITIGITIGIDAS 172  
QY 174 HPDLKRVIGWYDAVNGRSTPYDDQGHGTHVAGIVAGTGSV--NSQYIGVAPGAKLVGVK 232  
DB 173 HPDLQCKVIGWDFVNGRSYPRYDDHGHGTHVASIAAGTGAASNGKGMAPKAKLAGIKV 232  
QY 233 LGADSGSVSTIIAGVDWVYQNKDKYIRVINSLSGSSQSDGTDSLQAANNAMDAIV 292  
DB 233 LGADSGSVSTIIKGVEMAVDNKDKYIKVINSLSGSSQSDGTDSLQAANNAMDAIV 292  
QY 293 VCVAAAGNSGPNYTVGSPAAASKVITVGAVDSDNITASFSRGPRTADGRKPEVVA PGVD 352  
DB 293 VCVAAAGNSGPNYTVGSPAAASKVITVGAVDSDNITASFSRGPRTADGRKPEVVA PGVD 352  
QY 353 IAPRASGTSMTGPIINDYTTKASGTSMATPHVSGVALILQAHPSWTPDKVKTALJETAD 412  
DB 353 IAPRASGTSMTGPIINDYTTKASGTSMATPHVSGVALILQAHPSWTPDKVKTALJETAD 412  
QY 413 IYAPKEIADIAYAGRVNYYKAIKTDYAKLTFTGSVADKSGATHTFDVSGATFTATLY 472  
DB 413 IYAPKEIADIAYAGRVNYYKAIKTDYAKLTFTGSVADKSGATHTFDVSGATFTATLY 472  
QY 473 WPDGSSDIDLVLDPNGNEVDVSYTAAYGPEKGYGNPNTAGTWTVVVVSXGKAANYOVV 532  
DB 473 WPDGSSDIDLVLDPNGNEVDVSYTAAYGPEKGYGNPNTAGTWTVVVVSXGKAANYOVV 532  
QY 533 VSDGSLISOGSGGNPNPNPNPTPTDTOTFTGSVANDYMDTSDTFMANNVSGATKTIGDL 592  
DB 533 VSDGSLISOGSGGNPNPNPNPTPTDTOTFTGSVANDYMDTSDTFMANNVSGATKTIGDL 592  
QY 593 TPTSTINDLDLYYDRNGNLVDRSTGSNSYEHVEYANPAFGTWTPLVAYSTYGMADYQI 652  
DB 593 TPTSTINDLDLYYDRNGNLVDRSTGSNSYEHVEYANPAFGTWTPLVAYSTYGMADYQI 652  
QY 653 KAVVYTG 659  
DB 653 KAVVYTG 659

RESULT 7  
US-08-894-818B-35

Sequence 35, Application US/08894818B  
Patent No. 6261822  
GENERAL INFORMATION:  
APPLICANT: TAKAKURA, Hikaru  
APPLICANT: MORISHITA, Mio  
APPLICANT: YAMAMOTO, Katsuhiko  
APPLICANT: MITTA, Masamori  
APPLICANT: ASADA, Kiyozo  
APPLICANT: TSUNASAWA, Susumu  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: Browdy and Neimark  
STREET: 419 Seventh Street N.W., Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/894,818B  
FILING DATE: 20-MAY-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP96/03253  
FILING DATE: 07-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 323285/1995  
FILING DATE: 12-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TAKAKURA-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 654 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-894-818B-35

Query Match 72.3%; Score 2483.5; DB 2; Length 654;  
Best Local Similarity 72.1%; Pred. No. 1.4e-168;  
Matches 481; Conservative 69; Mismatches 96; Indels 21; Gaps 8;

QY 1 MRRLGAVTLALVGLLALTAAPKPV--VRNNAVQOKNYGLTPGLFKYQRMNMQ 58  
DB 1 MGKLKALILVILVLGLVGSVAAPKRYQVRN--VERKYGILTPGLFRKIQKLNPE 57  
QY 59 EVDVTIMFGSYGDRRAVAVLRMGAVKYSYKIIIPAVNKKIRADLLIAGMIDGYRG 118  
DB 58 EISTVIVFENHREKEIAVRVLELMGAKRVYVHIIPALADLKVRDLVLSGL--TG--G 113  
QY 119 NTRVSGIKFIQEDYKQVDDA-----TSVSIQADTVWNSLGYDSGVAIVDTGIDAN 173  
DB 114 KAKLSGVRFIQEYKXTVSALEGLDESAQVWATYVN--LGYDSGIGITIGITIGIDAS 172  
QY 174 HPDLKRVIGWYDAVNGRSTPYDDQGHGTHVAGIVAGTGSV--NSQYIGVAPGAKLVGVK 232  
DB 173 HPDLQCKVIGWDFVNGRSYPRYDDHGHGTHVASIAAGTGAASNGKGMAPKAKLAGIV 232  
QY 233 LGADSGSVSTIIAGVDWVYQNKDKYIRVINSLSGSSQSDGTDSLQAANNAMDAIV 292  
DB 233 LGADSGSVSTIIKGVEMAVDNKDKYIKVINSLSGSSQSDGTDSLQAANNAMDAIV 292

|     |     |  |     |
|-----|-----|--|-----|
| QY  | 293 | CVNAGNGSPNYYTYSGPSAAASKRTITVGAVDNSNDINLASFSSRQPTADGRUKPEVPA PGVD | 352 |
| Db  | 293 | VVVAAGNSGPNKTYTIGSPAAASKVITVGAVDKDYLTSSFSRQPTADGRUKPEVPA PGVN    | 352 |
| QY  | 353 | IIPASGSGTSMQPIINDYYTTKASGTSMAFPHVSGVALLILOAPSMTPDKEVKALLETAD     | 412 |
| Db  | 353 | IITAAASGSGMQPIINDYYTTAAGTSMATFPHVAGIALALLLOAPSMTPDKEVKALLETAD    | 412 |
| QY  | 413 | IIVAPKEIADIAYAGRVNYYKAIKYDDYAKLFTTGSVADKGSATHTTFDVSQATFTATLY     | 472 |
| Db  | 413 | IYKPEIADIAYAGRVNYYKAIINYNYAKLVFTGIVANKGSGTQHFVISGAFVATATLY       | 472 |
| .QY | 473 | MDTGSSDIDLXYLYDPNGNEVDVSYTYLYVPEKKGYYNPAAGTMTVTVSVKGAANTQVDY     | 532 |
| Db  | 473 | MDNANSDIDLXYLYDPNGNQVDYSYLYVPEKKGYYNPTDGTMTIKVSVSGSANTQVDY       | 532 |
| QY  | 533 | VSDGSLSGCGGNGPNPNPNPTPTDTQTFTGSVNDYMDTSDFTMTNVNSGATKITGDL        | 592 |
| Db  | 533 | VSDGSLSGQ-----PGSSPSBPQPEPTYDAKTFQSGSHYYIDRSITFTMTNVNSGATKITGDL  | 587 |
| QY  | 593 | TFDTSYNDLIDLXYLYDPNGNIVDRSTSSNSYEHEVYANPABGTMTFLVAYASTYGMADYOL   | 652 |
| Db  | 588 | VFDTSYHDLIDLXYLYDPNQKLVDRESPSNSYEHEVYLTLPAGTWYFLVAYASTYGMAYEL    | 647 |
| QY  | 653 | KAVVYTYG 659   |     |
| Db  | 648 | TAKVYTG 654  |     |

RESULT 8  
US-09-445-472-16  
; Sequence 16, Application US/09445472

Patent No.6358726  
GENERAL INFORMATION:  
APPLICANT: TAKAGURA, Hikaru  
APPLICANT: MORISHITA, MIO  
APPLICANT: SHIMOTO, Tomoko  
APPLICANT: ASADA, Kiyozo  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
FILE REFERENCE: TAKAGURA-6  
CURRENT APPLICATION NUMBER: US/09/445,472  
CURRENT FILING DATE: 1999-12-06  
PRIOR APPLICATION NUMBER: 151969/1997  
PRIOR FILING DATE: 1997-06-10  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 16  
LENGTH: 654  
TYPE: PRT  
ORGANISM: *Pyrococcus furiosus*  
US-09-445-472-16

|                           |        |                     |            |             |
|---------------------------|--------|---------------------|------------|-------------|
| Query Match               | 72.3%  | Score 2483.5;       | DB 2;      | Length 654; |
| Best Local Similarity     | 72.1%; | Pred. No. 1.4e-168; |            |             |
| Matches 481; Conservative | 69;    | No. Matches 96;     | Indels 21; | Gaps 8;     |

Qy 1 MMR1AAYVLAIVLVGLAGTALALAPKPV--PNNNAVOQKNTGLLPRGLFKKYQNNMNO 58  
Db 1 MGLKRLILITVLVLGVLGVSVAAPAKKVBQVNA--VEKNTGLLPPGLRFRK1QKLNPNB 57  
Qy 59 EYDVTVMFGSGYGRDPAVKVLRMLMGQOVKYSYKIIPAVAVKIKARDLLIAGMIDTGYG 118  
Db 58 ELSITYVFEENHREKELAAVRLVELMGAKRYVTHIIPALADLKVRLDLVLSGL--TG--G 113  
Qy 119 NTRVSGIKETIOBDYKQVDDA-----TSVQIGADTVMNLSLGYDGSQVVAIVDTGIDAN 173  
Db 114 KAKLSGVRFIQBDYKVTYSALEGLGDEBSAAQVAVATVMN--LGVDSGIGITGIDTGDAS 172  
Qy 174 HPDLKRVIGWYDVAVNGRSTPYDDQGHGTHVAGIVAGTGSV--NSQYITGVAAPKALVGVKY 232  
Db 173 HPDLQKRVIGWYDVNGRSTPYDDHGHGTHVASIAGTGAASNGKYGMAFGKLAGIITY 232

|    |     |  |     |
|----|-----|--|-----|
| OY | 233 | LGAOGSGSVSTI IAGVDWVVOXKDKYGRIVINLSIGSSQOSSPGDTBSLQAANNAMDA GIV    | 292 |
| Dd | 233 | LGAOBSGSISTIIKGVEMAVDNKKTKGIKVINLSIGSSQOSSDGTDLAQVAAMDAGIV         | 292 |
| OY | 293 | VCAAGNSGPNVTYTTSVPAAASKVITTVGAVDSNDNIASFSSHQPTADGLKCBVAPGD         | 352 |
| Dd | 293 | VYVAAGNSGPNKYRTTIGSPAAASKVITTVGAVDYTDYITSFSSRGPTADGLKCBVAPGMW      | 352 |
| OY | 353 | IIPPAASGTSNGTPINDYYTTASGTSMACTPHVSGVALILOHPBMTPKXYTALLETAD         | 412 |
| Dd | 353 | IIAAASGTSNGQPIINDYYTAAPGTSMATPHVAGIALLOLHPBMTPKXYTALLETAD          | 412 |
| OY | 413 | IVAPKEIADIAYVAGRVANVYKAIKYDDPAKLFTFSVADKCSATHFDVSAGATEVATILY       | 472 |
| Dd | 413 | IYKPBEIADIAYGAGRVANAYKANINDYNPAKLFTFYVANAKGSQTHQFVISGASFVATILY     | 472 |
| OY | 473 | MDTGSSDIDLTYLPDNGNEBVDTSYTAAYGGFEKVGYNPPTAGTWYKVVYSYKGAANYQUDV     | 532 |
| Dd | 473 | WDNAASDDLTYLPDNGNQUDVSYTAAYGFEEKGYNNPPTDGXTWIKVYSYSASANYQUDV       | 532 |
| OY | 533 | VSDGSLSGSGGNPNPNPNPTPTPDQTFFGGSVNDVWDTSDTFPMNNSGATKITTGD           | 592 |
| Dd | 533 | VSDGSLSG- - - - - PESSPSPODEPVVDAKTFFGGSDBHYIDRKSDTFTMYNSGATKITTGD | 597 |
| OY | 593 | TFTDTSYNDLDLYLPDPNGNLVDRTSSNSYEHEVBYANBPAGTWTFLVYASYTYMADYOL       | 652 |
| Dd | 588 | VFDTSYHDLIDLTYLPDPNQKLVDRSSSPNSYEHEVELTPAPGTWYFLVAYATYTMAYEEL      | 647 |
| OY | 653 | KAVVYTG 659  |     |
| Dd | 648 | TAKVYTG 654  |     |

RESULT 9  
US-10-090-624-16  
; Sequence 16, Application US/10090624

```

/ Patent No.6783970
/ GENERAL INFORMATION:
/ APPLICANT: TAKAKURA, Hikaru
/ APPLICANT: MORISHITA, Mio
/ APPLICANT: SHIMODO, Tomoko
/ APPLICANT: ASADA, Kiyozo
/ APPLICANT: KATO, Ikunoshin
/ TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
/ FILE REFERENCE: TAKAKURA-6
/ CURRENT APPLICATION NUMBER: US/10/090,624
/ CURRENT FILING DATE: 2002-03-06
/ PRIOR APPLICATION NUMBER: 09/445,472
/ PRIOR FILING DATE: 1999-12-06
/ PRIOR APPLICATION NUMBER: 151969/1997
/ PRIOR FILING DATE: 1997-06-10
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 16
/ LENGTH: 654
/ TYPE: PRT
/ ORGANISM: Pyrococcus furiosus
/
US-10-090-624-16

```

|    |                       |   |  |                |                    |
|----|-----------------------|---|--|----------------|--------------------|
|    | Query Match           | 72.3%                                   | Score 2483.5;  | DB 2;          | Length 654;        |
|    | Best Local Similarity | 72.1%;                                  | Pred. No. 1.4e-168;  |                |                    |
|    | Matches               | 481;                                    | Conservative 69;   | Mismatches 96; | Indels 21; Gaps 8; |
| Qy |                       | 1                                       | MKRIGAVVLALVILGLAGTALAAPVKVV--VRANNAWQQKNYGLTLTPGLFKFVKVORMNWQ | 58             |                    |
|    |                       | :   :   :   :                 :   :   : |  |                |                    |
| Dd |                       | 1                                       | MKGLKALLLVILVILGLVGVAAPPEKKVEOVRN---VEKYTGILLTPGLFKIKLQNPNE    | 57             |                    |
| Qy |                       | 59                                      | EVDVTIMFGSGDDRDRAVKVRLMGACQYKYSYKIIIPAAVXIKARDDLILGMIIDTYEGF   | 118            |                    |
|    |                       | :   :   :   :                 :   :   : |  |                |                    |
| Dd |                       | 58                                      | EISLYIYEENREKEINAVRVELMGAKRYYYHIIIPALADLKVRDLLVISGL--TG--G     | 113            |                    |
| Qy |                       | 119                                     | NTRVSQIFIQEDBKVQVDDA-----TSVSOIGADITWNLSIGDGSGVVAIVDTGDIDAN    | 173            |                    |

```

Db      114 KAKLSGVRFIQEDYKVTVALEGLDESAQVMAVYVWN-LGVDGSGITIGITDGDAS 172
Qy      174 HPDLKGVIGWVAVNVRSTPYDDQGHGTHVAGI VAGTGSV-NSQYIYVAPGAKLVGVY 232
Db      173 HPDLQGVIGWVAVNVRSTPYDDHGHGTHVAGI IAGTGAASNGKYKGMAPGAKLVGIV 232
Qy      233 LGADGSGSVSTIAGVDMVONKDKYGI RIVINLSLSSGSSDGTDSLQAVNNAMPAGIV 232
Db      233 LGADGSGSVSTIAGVDMVONKDKYGI RIVINLSLSSGSSDGTDSLQAVNNAMPAGIV 232
Qy      293 VCVAAAGNSGPNYTVGSSPAASKVITVGA VDSNDNIASFSSRGPPTADGRLPKPEVAPGV 352
Db      293 VVVAAGNSGPNKTTISGPAASKVITVGA VDKDVLTSFSSRGPPTADGRLPKPEVAPGV 352
Qy      353 IIAPRASGTMGPINDYTYKASGTSMA TPVHVGALLIQAHPSWTPDKVKTALJETAD 412
Db      353 IIAPRASGTMGPINDYTYKASGTSMA TPVHVGALLIQAHPSWTPDKVKTALJETAD 412
Qy      413 IYAPKEIADIAYAGAGRVNYYKAIKYD DYAFLTFTGSVADKSGATHTFPDVSAGATEVATLY 472
Db      413 IYAPKEIADIAYAGAGRVNYYKAIKYD DYAFLTFTGSVADKSGATHTFPDVSAGATEVATLY 472
Qy      473 WDTGSSDIDLXYDPNGNEVDYSYTA YGFEKGYNPNPTAGTWTVKVVSYKGAANTQVDY 532
Db      473 WDNANSDLDLYDPNGNQVDYSYTA YGFEKGYNPNPTAGTWTIKVVSYSGSANTQVDY 532
Qy      533 VSDGSLSSQSGGPNPNPNPTPTD TQFTGSVNDYWDTSPTFMNNSGATKTITGDL 552
Db      533 VSDGSLSSQSGGPNPNPNPTPTD TQFTGSVNDYWDTSPTFMNNSGATKTITGDL 552
Qy      593 TPDTSNDLDLYDPNGNLYDRSST SSVSEHYEVANPAPGTWPLVYASTYGMADYOL 652
Db      593 TPDTSNDLDLYDPNGNLYDRSST SSVSEHYEVANPAPGTWPLVYASTYGMADYOL 652
Qy      653 KAVVYVG 659
Db      648 TAKVYVG 654

```

RESULT 10  
US-09-841-553-35  
Sequence 35, Application US/09841553  
Patent No. 6843441  
GENERAL INFORMATION:  
APPLICANT: TAKAKURA, Hikaru  
MORISHITA, Mio  
YAMAMOTO, Katsuhiko  
MITTA, Masanori  
ASADA, Kiyozo  
TSUNASAWA, Susumu  
KATO, Ikunoshin  
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street N.W., Ste. 300  
City: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/841,553  
FILING DATE: 24-Apr-2001  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/894,818  
FILING DATE: <Unknown>  
APPLICATION NUMBER: JP 323285/1995

```

Filing Date: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 654 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-09-841-553-35

Query Match      72.3%; Score 2483.5; DB 2; Length 654;
Best Local Similarity 72.1%; Pred. No. 1.4e-168;
Matches 481; Conservative 69; Mismatches 96; Indels 21; Gaps 8;

1 MKRLGAVLALVVLGSLAGTALAAPYKPV--VRNNAVQQRNTGLTTPGLFKYVQRNMMQ 58
1 MGGLKALLIIVILVGLHVGSAVAAPPEKKEOVARN--VEKNYGLLTPGLFRKIQKLNPN 57
59 EVDVTIMFGSYGRDRDAVAVVRLMGAVKYKTIIPAVAVKIKARPLLIIAGMIDGYRG 118
58 EISTVIVFNHRKELIAVAVLELMGAKVRYVHIIPALIAADLKRDLVLSGL--TG--G 113
119 NTRVSGIKFIQEDYKQVQVDA-----TSVQICADTVNMSLGVDGSGVVAIVDTGIDAN 173
114 KAKLSGVRFIQEDYKVTVALEGLDESAQVMAVYVWN-LGVDGSGITIGITDGDAS 172
174 HPDLKGVIGWVAVNVRSTPYDDQGHGTHVAGI VAGTGSV-NSQYIYVAPGAKLVGVY 232
173 HPDLQGVIGWVAVNVRSTPYDDHGHGTHVAGI IAGTGAASNGKYKGMAPGAKLVGIV 232
233 LGADGSGSVSTIAGVDMVONKDKYGI RIVINLSLSSGSSDGTDSLQAVNNAMPAGIV 292
233 LGADGSGSVSTIAGVDMVONKDKYGI RIVINLSLSSGSSDGTDSLQAVNNAMPAGIV 292
293 VCVAAAGNSGPNYTVGSSPAASKVITVGA VDSNDNIASFSSRGPPTADGRLPKPEVAPGV 352
293 VVVAAGNSGPNKTTISGPAASKVITVGA VDKDVLTSFSSRGPPTADGRLPKPEVAPGV 352
353 IIAPRASGTMGPINDYTYKASGTSMA TPVHVGALLIQAHPSWTPDKVKTALJETAD 412
353 IIAPRASGTMGPINDYTYKASGTSMA TPVHVGALLIQAHPSWTPDKVKTALJETAD 412
413 IYAPKEIADIAYAGAGRVNYYKAIKYD DYAFLTFTGSVADKSGATHTFPDVSAGATEVATLY 472
413 IYAPKEIADIAYAGAGRVNYYKAIKYD DYAFLTFTGSVADKSGATHTFPDVSAGATEVATLY 472
473 WDTGSSDIDLXYDPNGNEVDYSYTA YGFEKGYNPNPTAGTWTVKVVSYKGAANTQVDY 532
473 WDNANSDLDLYDPNGNQVDYSYTA YGFEKGYNPNPTAGTWTIKVVSYSGSANTQVDY 532
533 VSDGSLSSQSGGPNPNPNPTPTD TQFTGSVNDYWDTSPTFMNNSGATKTITGDL 592
533 VSDGSLSSQSGGPNPNPNPTPTD TQFTGSVNDYWDTSPTFMNNSGATKTITGDL 592
593 TPDTSNDLDLYDPNGNLYDRSST SSVSEHYEVANPAPGTWPLVYASTYGMADYOL 652
593 TPDTSNDLDLYDPNGNLYDRSST SSVSEHYEVANPAPGTWPLVYASTYGMADYOL 652
653 KAVVYVG 659
648 TAKVYVG 654

```

RESULT 11  
US-08-894-818B-3

```

; Sequence 3, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 737-3528
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; OTHER INFORMATION: /note=Xaa at position 428 is Gly or Val.
; US-08-894-818B-3

Query Match 62.2%; Score 2138.5; DB 2; Length 522;
Best Local Similarity 78.3%; Pred. No. 3.9e-144;
Matches 407; Conservative 36; Mismatches 70; Indels 7; Gaps 3;

QY 141 SVQIGADVTWNSLIGDGSVVVAIVDTGIDANHPDLKGVIGWYDAVNGRSTPYDDOGH 200
DB 9 SAAQVATYVWN-LGYDSSGITIGIIDTGIDASHPDLQKVGWDPVNGRSTPYDDHGH 67

QY 201 GTHVAGIVAGTGSV-NSQYIGVAPGAKLVGVKVLGADGSGSVSTIIAGVDMVWVONKDKYG 259
DB 68 GTHVASIAAGTGAASNGKYGMAPGAOKLAGIKVLGADGSGSISTIIKGVEMAVDNKDKYG 127

QY 260 IRVINSLGSSQSSDGTDSLQAANNAMDAIGIVCVAAAGNSGPNYTVGSPAAASKVITV 319
DB 128 IKVINSLGSSQSSDGTDSLQAANNAMDAIGIVCVAAAGNSGPNKTTIGSPAAASKVITV 187

QY 320 GAVDSNDNIAFSRSGRTADGRLKPEVAPGVVDIIAPRASGTMGTPINDYTTKASGTM 379
DB 188 GAVDKYDVITFSRSGRTADGRLKPEVAPGVNMIIAARASGTMGQPINDYTTAAGTSM 247

QY 380 ATPHVSQVALLIQAPHSWTPDKVKTALIBTADIVAPKEIADIAYGAGRNVYKAIKYDD 439

```

```

DB 248 ATPHVAGIAALLIQAPHSWTPDKVKTALIBTADIVAPKEIADIAYGAGRNVAKAIINYDN 307
QY 440 YAKLFTGSAVDGSAATHFTDVSGATFTVATTLTYMDGSSPIDLYLDPNGNENEDYSTAY 439
DB 308 YAKLFTGVVANGSGQTHQFVIGASFTVATTLWMDANSDLDLYLDPNGNOVDYSTAY 367
QY 500 YGFEKGYVNPATGATTVKVASYKGAANYQVDVVSQSLSGSGGPNPNPNPTPTD 559
DB 368 YGFEKGYVNPATGATTVKVASYKGAANYQVDVVSQSLSGSGGPNPNPNPTPTD 422
QY 560 TQFTFGSVNDYMDTSPFTFNNVNSGATKITGDTFTPTSYNDLIDLYLDPNGNLVDRSTSS 619
DB 423 AKTFQXSDHYVYRSDFTFTVNSGATKITGDLVFTDSYHDLIDLYLDPNGNLVDRSESP 482
QY 620 NSYEHVEYANPARGTWTFLVVAYSTGMDYOLKAVVYVG 659
DB 483 NSYEHVEYLTAPGTWTFVLVAYTYVGMAYEILTAVYVG 522

RESULT 12
US-09-445-472-4
; Sequence 4, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (428)..(428)
; OTHER INFORMATION: Xaa at position 428 is Gly or Val.
; US-09-445-472-4

Query Match 62.2%; Score 2138.5; DB 2; Length 522;
Best Local Similarity 78.3%; Pred. No. 3.9e-144;
Matches 407; Conservative 36; Mismatches 70; Indels 7; Gaps 3;

QY 141 SVQIGADVTWNSLIGDGSVVVAIVDTGIDANHPDLKGVIGWYDAVNGRSTPYDDOGH 200
DB 9 SAAQVATYVWN-LGYDSSGITIGIIDTGIDASHPDLQKVGWDPVNGRSTPYDDHGH 67

QY 201 GTHVAGIVAGTGSV-NSQYIGVAPGAKLVGVKVLGADGSGSVSTIIAGVDMVWVONKDKYG 259
DB 68 GTHVASIAAGTGAASNGKYGMAPGAOKLAGIKVLGADGSGSISTIIKGVEMAVDNKDKYG 127

QY 260 IRVINSLGSSQSSDGTDSLQAANNAMDAIGIVCVAAAGNSGPNYTVGSPAAASKVITV 319
DB 128 IKVINSLGSSQSSDGTDSLQAANNAMDAIGIVCVAAAGNSGPNKTTIGSPAAASKVITV 187

QY 320 GAVDSNDNIAFSRSGRTADGRLKPEVAPGVVDIIAPRASGTMGTPINDYTTKASGTM 379
DB 188 GAVDKYDVITFSRSGRTADGRLKPEVAPGVNMIIAARASGTMGQPINDYTTAAGTSM 247

QY 380 ATPHVSQVALLIQAPHSWTPDKVKTALIBTADIVAPKEIADIAYGAGRNVYKAIKYDD 439
DB 248 ATPHVAGIAALLIQAPHSWTPDKVKTALIBTADIVAPKEIADIAYGAGRNVYKAIINYDN 307

QY 440 YAKLFTGSAVDGSAATHFTDVSGATFTVATTLTYMDGSSPIDLYLDPNGNENEDYSTAY 439

```





```

Db      9 SAQVMAATYWN-LGYDGSGITIGIDTGDASHPDLQGVKIVGVDVFNGRSTPYDDHG 67
Qy      201 GTHVAGIYAATGSV-NSQYIGVAPGATLVGVKVLGADGSSVSTIINGVWVYQNDKYG 259
Db      68 GTHVASTIAACTGAASNGKYGMAPGATLAIKVLGAGSGSISTIIKGVMAVDNKKYK 127
Qy      260 IRVINSLSSQSSDGTDSLQAVNNMADAGIYVCAAGNSGPNYTVGSPAASKYITV 319
Db      128 IKVINSLSSQSSDGTDSLQAVNNMADAGIYVCAAGNSGPNYTVGSPAASKYITV 187
Qy      320 GAVDSNDNIASFSSRGPTADGRLKPEVVAPEGVDIIAPRASGSMGPIINDYTTKASGTM 379
Db      188 GAVDKYDVITSFSSRGPTADGRLKPEVVAPEGVDIIAPRASGSMGPIINDYTTKASGTM 247
Qy      380 ATPHVSQVGLIIQAHPSWTPDKVKTALITFADIAPKEIADIAYGAGRVNYYKAIKYDD 439
Db      248 ATPHVAGIALLIQAHPSWTPDKVKTALITFADIAPKEIADIAYGAGRVNYYKAIKYDD 307
Qy      440 YAKLFTGSVADKGSATHTFDVSGATFVATLYMTDSSDIDLYLDPNGNEVDYSYAY 499
Db      308 YALVFTGYVANKGSQTHQFVIGSASFVATLYWDMANSDLDLYLDPNGNQVDYSYAY 367
Qy      500 YGFEKGYNNPTAGTWTVKVSYKGAANYQVDVSDGSLSGGQGNPNPNPTPTD 559
Db      368 YGFEKGYNNPTDGTWTIKVSYSGSANYQVDVSDGSLSG-----FGSSSPQPEPTVD 422
Qy      560 TQFTGSVNDYMTDPTFETMNVNNGSAGTKITGDLTFPDSYNDLDLYLDPNGNVDSTSS 619
Db      423 AKTFQASDHYTDSDFTFTVNSGATKITGDLVFDTSYHDLIDLYLDPNGKLVDSRSP 482
Qy      620 NSYEHVEYANPAPGTWTLVYAYSTYGMADYOLKAVVYVG 659
Db      483 NSYEHVEYLTLPAPGTWTLVYAYTYGMAYYBLAKVYVG 522

```

```

RESULT 15
US-09-445-472-1
; Sequence 1, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-09-445-472-1

```

```

Query Match      49.7%; Score 1707; DB 2; Length 412;
Best Local Similarity 80.6%; Pred. No. 1.se-113;
Matches 325; Conservative 29; Mismatches 47; Indels 2; Gaps 2;

```

```

Qy      141 SVSOIGADTYWNSLGYDGSVVAIYDTGIDANHPDLKGVIGVDAVNGRSTPYDDQH 200
Db      9 SAQVMAATYWN-LGYDGSGITIGIDTGDASHPDLQGVKIVGVDVFNGRSTPYDDHG 67
Qy      201 GTHVAGIYAATGSV-NSQYIGVAPGATLVGVKVLGADGSSVSTIINGVWVYQNDKYG 259
Db      68 GTHVASTIAACTGAASNGKYGMAPGATLAIKVLGAGSGSISTIIKGVMAVDNKKYK 127
Qy      260 IRVINSLSSQSSDGTDSLQAVNNMADAGIYVCAAGNSGPNYTVGSPAASKYITV 319

```

```

Db      128 IKVINSLSSQSSDGTDSLQAVNNMADAGIYVCAAGNSGPNYTVGSPAASKYITV 187
Qy      320 GAVDSNDNIASFSSRGPTADGRLKPEVVAPEGVDIIAPRASGSMGPIINDYTTKASGTM 379
Db      188 GAVDKYDVITSFSSRGPTADGRLKPEVVAPEGVDIIAPRASGSMGPIINDYTTKASGTM 247
Qy      380 ATPHVSQVGLIIQAHPSWTPDKVKTALITFADIAPKEIADIAYGAGRVNYYKAIKYDD 439
Db      248 ATPHVAGIALLIQAHPSWTPDKVKTALITFADIAPKEIADIAYGAGRVNYYKAIKYDD 307
Qy      440 YAKLFTGSVADKGSATHTFDVSGATFVATLYMTDSSDIDLYLDPNGNEVDYSYAY 499
Db      308 YALVFTGYVANKGSQTHQFVIGSASFVATLYWDMANSDLDLYLDPNGNQVDYSYAY 367
Qy      500 YGFEKGYNNPTAGTWTVKVSYKGAANYQVDVSDGSLSGGQGNPNPNPTPTD 542
Db      368 YGFEKGYNNPTDGTWTIKVSYSGSANYQVDVSDGSLSGQPS 410

```

Search completed: January 6, 2007, 22:33:39  
Job time : 37 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2007 Bioacceleration Ltd.

OM protein - protein search, using SW model

Run on: January 6, 2007, 22:31:42 ; Search time 36 Seconds  
(without alignments)  
1602.297 Million cell updates/sec

Title: US-10-800-684-5

Perfect score: 3428  
Sequence: 1 MKGLKALILVILVGLHVG.....YAVSTYGMADYOLKAVVYVG 659

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: Issued Patents NA:  
2: /EMC\_Celerra\_SIDS3/Pctodata/2/1aa/5/COMB.pep:\*  
3: /EMC\_Celerra\_SIDS3/Pctodata/2/1aa/6/COMB.pep:\*  
4: /EMC\_Celerra\_SIDS3/Pctodata/2/1aa/7/COMB.pep:\*  
5: /EMC\_Celerra\_SIDS3/Pctodata/2/1aa/H/COMB.pep:\*  
6: /EMC\_Celerra\_SIDS3/Pctodata/2/1aa/PCTUS.COMB.pep:\*  
7: /EMC\_Celerra\_SIDS3/Pctodata/2/1aa/backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score  | Query Match | Length DB | ID | Description       |
|------------|--------|-------------|-----------|----|-------------------|
| 1          | 3428   | 100.0       | 659       | 2  | US-08-894-818B-5  |
| 2          | 3428   | 100.0       | 659       | 2  | US-09-841-553-5   |
| 3          | 2997.5 | 87.4        | 654       | 2  | US-08-894-818B-35 |
| 4          | 2997.5 | 87.4        | 654       | 2  | US-09-445-472-16  |
| 5          | 2997.5 | 87.4        | 654       | 2  | US-10-090-624-16  |
| 6          | 2997.5 | 87.4        | 654       | 2  | US-09-841-553-35  |
| 7          | 2997.5 | 87.4        | 659       | 2  | US-08-894-818B-1  |
| 8          | 2997.5 | 87.4        | 659       | 2  | US-09-445-472-12  |
| 9          | 2997.5 | 87.4        | 659       | 2  | US-10-090-624-12  |
| 10         | 2997.5 | 87.4        | 659       | 2  | US-09-841-553-1   |
| 11         | 2997.5 | 87.4        | 659       | 2  | US-08-894-818B-3  |
| 12         | 2997.5 | 87.4        | 659       | 2  | US-09-445-472-4   |
| 13         | 2997.5 | 87.4        | 659       | 2  | US-10-090-624-4   |
| 14         | 2997.5 | 87.4        | 659       | 2  | US-09-841-553-3   |
| 15         | 2997.5 | 87.4        | 659       | 2  | US-09-445-472-1   |
| 16         | 2997.5 | 87.4        | 659       | 2  | US-10-090-624-1   |
| 17         | 2997.5 | 87.4        | 659       | 2  | US-09-841-553-1   |
| 18         | 2997.5 | 87.4        | 659       | 2  | US-08-894-818B-3  |
| 19         | 2997.5 | 87.4        | 659       | 2  | US-09-445-472-4   |
| 20         | 2997.5 | 87.4        | 659       | 2  | US-10-090-624-4   |
| 21         | 2997.5 | 87.4        | 659       | 2  | US-09-841-553-3   |
| 22         | 2997.5 | 87.4        | 659       | 2  | US-09-445-472-1   |
| 23         | 2997.5 | 87.4        | 659       | 2  | US-10-090-624-1   |
| 24         | 2997.5 | 87.4        | 659       | 2  | US-09-841-553-1   |
| 25         | 2997.5 | 87.4        | 659       | 2  | US-08-894-818B-3  |
| 26         | 2997.5 | 87.4        | 659       | 2  | US-09-445-472-4   |

|    |       |      |     |   |                  |                    |
|----|-------|------|-----|---|------------------|--------------------|
| 27 | 555   | 16.2 | 382 | 2 | US-10-104-693-2  | Sequence 2, Appl1  |
| 28 | 554.5 | 16.2 | 378 | 1 | US-07-772-087-4  | Sequence 4, Appl1  |
| 29 | 554   | 16.2 | 379 | 1 | US-08-845-295A-1 | Sequence 1, Appl1  |
| 30 | 554   | 16.2 | 379 | 2 | US-09-140-933-1  | Sequence 1, Appl1  |
| 31 | 554   | 16.2 | 379 | 2 | US-09-146-661-1  | Sequence 1, Appl1  |
| 32 | 554   | 16.2 | 379 | 2 | US-09-150-515-1  | Sequence 1, Appl1  |
| 33 | 551   | 16.1 | 382 | 2 | US-09-255-502-2  | Sequence 2, Appl1  |
| 34 | 551   | 16.1 | 382 | 2 | US-09-659-749-2  | Sequence 2, Appl1  |
| 35 | 551   | 16.1 | 382 | 2 | US-09-672-105-2  | Sequence 2, Appl1  |
| 36 | 550   | 16.0 | 382 | 7 | 5472855-2        | Patent No. 5472855 |
| 37 | 549.5 | 16.0 | 381 | 1 | US-07-772-087-6  | Sequence 6, Appl1  |
| 38 | 549.5 | 16.0 | 381 | 1 | US-08-173-508-12 | Sequence 12, Appl1 |
| 39 | 549.5 | 16.0 | 381 | 1 | US-08-265-310-12 | Sequence 12, Appl1 |
| 40 | 549.5 | 16.0 | 381 | 2 | US-08-951-742-12 | Sequence 2, Appl1  |
| 41 | 547   | 16.0 | 382 | 2 | US-09-178-155-2  | Sequence 2, Appl1  |
| 42 | 547   | 16.0 | 382 | 2 | US-09-178-173A-2 | Sequence 2, Appl1  |
| 43 | 547   | 16.0 | 382 | 2 | US-09-177-353-2  | Sequence 2, Appl1  |
| 44 | 547   | 16.0 | 382 | 2 | US-10-033-325-2  | Sequence 2, Appl1  |
| 45 | 547   | 16.0 | 382 | 2 | US-10-828-572-2  | Sequence 2, Appl1  |

## ALIGNMENTS

RESULT 1  
US-08-894-818B-5  
Sequence 5, Application US/08894818B  
Patent No. 6261822  
GENERAL INFORMATION:  
APPLICANT: TAKAKURA, Hikaru  
APPLICANT: MORISHITA, Mio  
APPLICANT: YAMAMOTO, Katsuhiko  
APPLICANT: MITTA, Masanori  
APPLICANT: ASADA, Kiyozo  
APPLICANT: TSUNASAKA, Susumu  
APPLICANT: KATO, Ikumoshin  
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: Browdy and Neimark  
STREET: 419 Seventh Street N.W., Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/894, 818B  
CLASSIFICATION: 435  
FILING DATE: 20-May-1998  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP96/03253  
FILING DATE: 07-Nov-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 323285/1995  
FILING DATE: 12-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TAKAKURA=1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 659 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

*Div. of this application*  
*This reference*  
*seq id no. 3 (p.p)*  
*seq id no. 4 (DNA)*  
*No db.pat. problems.*

MOLECULE TYPE: peptide  
US-08-894-818B-5

Query Match 100.0%; Score 3428; DB 2; Length 659;  
Best Local Similarity 100.0%; Pred. No. 3.9e-241;  
Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKGKALILVILVGLVGVSAAPKPKVQVNRVKNYGLTPGLFRKIQKLNPNRIS 60  
DB 1 MKGKALILVILVGLVGVSAAPKPKVQVNRVKNYGLTPGLFRKIQKLNPNRIS 60  
QY TVIVFENHREKEIAVRLELMGAKRVYVHIIPALADLVKRDLLVSGLTGKAKLSGV 120  
DB TVIVFENHREKEIAVRLELMGAKRVYVHIIPALADLVKRDLLVSGLTGKAKLSGV 120  
QY 121 RFIQEDYKTVSAELRGLDESAQVMAVYVNIQYDGSITIGIIDTGDASHPDLQKV 180  
DB 121 RFIQEDYKTVSAELRGLDESAQVMAVYVNIQYDGSITIGIIDTGDASHPDLQKV 180  
QY 181 IGWDFVNGRSYPYDDHGHGTHVASTAAGTGAASNGKYGMAPGAAGKVLGADSGS 240  
DB 181 IGWDFVNGRSYPYDDHGHGTHVASTAAGTGAASNGKYGMAPGAAGKVLGADSGS 240  
QY 241 ISTTIKGVMAVNDKDKYGIKVINLSGSSQSDGTDLSQAVNNAMDAGIIVCVAAGNS 300  
DB 241 ISTTIKGVMAVNDKDKYGIKVINLSGSSQSDGTDLSQAVNNAMDAGIIVCVAAGNS 300  
QY 301 GPNITVYVGSPPAASKVITVGAVDSDNINIAFSSRGPADGRLEKPEVAVPGVDIIAPRASG 360  
DB 301 GPNITVYVGSPPAASKVITVGAVDSDNINIAFSSRGPADGRLEKPEVAVPGVDIIAPRASG 360  
QY 361 TSMGTPINDYTKASGTSMAATPHVSGVGLIIOAHPSWTPDKYKALIFRADIAPKEIA 420  
DB 361 TSMGTPINDYTKASGTSMAATPHVSGVGLIIOAHPSWTPDKYKALIFRADIAPKEIA 420  
QY 421 DIAYGARNVYVYAIKRYDDAKLFTFGSVADKGSATTFPVSAGATPTATILYMDTGSDDI 480  
DB 421 DIAYGARNVYVYAIKRYDDAKLFTFGSVADKGSATTFPVSAGATPTATILYMDTGSDDI 480  
QY 481 DLVLYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTVKVSYKGAANYQVDVSDGSLSQ 540  
DB 481 DLVLYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTVKVSYKGAANYQVDVSDGSLSQ 540  
QY 541 SGGGNPNPNPNPTPTDQTFTGTVNDYDTSDFTFMNVNSGATKITGDLTFDTSYND 600  
DB 541 SGGGNPNPNPNPTPTDQTFTGTVNDYDTSDFTFMNVNSGATKITGDLTFDTSYND 600  
QY 601 LDLYIDPNNGNLYDRSSNSYEHVEYANPAPGTWTFVLYASTYGMADYQLAAVYYG 659  
DB 601 LDLYIDPNNGNLYDRSSNSYEHVEYANPAPGTWTFVLYASTYGMADYQLAAVYYG 659

RESULT 2  
US-09-841-553-5

Sequence 5, Application US/09841553  
Patent No. 6849441

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru

MORISHITA, Mio

YAMAMOTO, Katsuhiko

MITTA, Masanori

ASADA, Kiyozo

TSUDASAWA, Susumu

KATO, Ikunoshin

TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300

CITY: Washington

STATE: D.C.

COUNTRY: United States of America

ZIP: 20004

*Handwritten notes:*  
of interest  
only Dist 5-27-12  
Dist 10-15-12  
N/A Dist 10-15-12

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/841,553

FILING DATE: 24-Apr-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/894,818

FILING DATE: <Unknown>

APPLICATION NUMBER: JP 323285/1995

FILING DATE: 12-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Browdy, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TAKAKURA-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 628-5197

TELEFAX: (202) 737-3528

INFORMATION FOR SEQ ID NO: 5

SEQUENCE CHARACTERISTICS:

LENGTH: 659 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-841-553-5

Query Match 100.0%; Score 3428; DB 2; Length 659;  
Best Local Similarity 100.0%; Pred. No. 3.9e-241;  
Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKGKALILVILVGLVGVSAAPKPKVQVNRVKNYGLTPGLFRKIQKLNPNRIS 60  
DB 1 MKGKALILVILVGLVGVSAAPKPKVQVNRVKNYGLTPGLFRKIQKLNPNRIS 60  
QY TVIVFENHREKEIAVRLELMGAKRVYVHIIPALADLVKRDLLVSGLTGKAKLSGV 120  
DB TVIVFENHREKEIAVRLELMGAKRVYVHIIPALADLVKRDLLVSGLTGKAKLSGV 120  
QY 121 RFIQEDYKTVSAELRGLDESAQVMAVYVNIQYDGSITIGIIDTGDASHPDLQKV 180  
DB 121 RFIQEDYKTVSAELRGLDESAQVMAVYVNIQYDGSITIGIIDTGDASHPDLQKV 180  
QY 181 IGWDFVNGRSYPYDDHGHGTHVASTAAGTGAASNGKYGMAPGAAGKVLGADSGS 240  
DB 181 IGWDFVNGRSYPYDDHGHGTHVASTAAGTGAASNGKYGMAPGAAGKVLGADSGS 240  
QY 241 ISTTIKGVMAVNDKDKYGIKVINLSGSSQSDGTDLSQAVNNAMDAGIIVCVAAGNS 300  
DB 241 ISTTIKGVMAVNDKDKYGIKVINLSGSSQSDGTDLSQAVNNAMDAGIIVCVAAGNS 300  
QY 301 GPNITVYVGSPPAASKVITVGAVDSDNINIAFSSRGPADGRLEKPEVAVPGVDIIAPRASG 360  
DB 301 GPNITVYVGSPPAASKVITVGAVDSDNINIAFSSRGPADGRLEKPEVAVPGVDIIAPRASG 360  
QY 361 TSMGTPINDYTKASGTSMAATPHVSGVGLIIOAHPSWTPDKYKALIFRADIAPKEIA 420  
DB 361 TSMGTPINDYTKASGTSMAATPHVSGVGLIIOAHPSWTPDKYKALIFRADIAPKEIA 420  
QY 421 DIAYGARNVYVYAIKRYDDAKLFTFGSVADKGSATTFPVSAGATPTATILYMDTGSDDI 480  
DB 421 DIAYGARNVYVYAIKRYDDAKLFTFGSVADKGSATTFPVSAGATPTATILYMDTGSDDI 480  
QY 481 DLVLYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTVKVSYKGAANYQVDVSDGSLSQ 540  
DB 481 DLVLYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTVKVSYKGAANYQVDVSDGSLSQ 540  
QY 541 SGGGNPNPNPNPTPTDQTFTGTVNDYDTSDFTFMNVNSGATKITGDLTFDTSYND 600  
DB 541 SGGGNPNPNPNPTPTDQTFTGTVNDYDTSDFTFMNVNSGATKITGDLTFDTSYND 600

QY 601 LDLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFELVYASTYGMADYQLKAVVYVG 659  
 DB 601 LDLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFELVYASTYGMADYQLKAVVYVG 659

## RESULT 3

US-08-894-818B-35  
 ; Sequence 35, Application US/088994818B  
 ; Patent No. 6261822  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TAKAKURA, Hikaru  
 ; APPLICANT: MORISHITA, Mio  
 ; APPLICANT: YAMAMOTO, Katsuhiko  
 ; APPLICANT: MITTA, Masanori  
 ; APPLICANT: ASADA, Kiyozo  
 ; APPLICANT: TSUNASAWA, Susumu  
 ; APPLICANT: KATO, Ikunoshin  
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
 ; NUMBER OF SEQUENCES: 42  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Browdy and Neimark  
 ; STREET: 419 Seventh Street N.W., Ste. 300  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: United States of America  
 ; ZIP: 20004  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/894,818B  
 ; FILING DATE: 20-MAY-1998  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/JP96/03253  
 ; FILING DATE: 07-NOV-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 323285/1995  
 ; FILING DATE: 12-DEC-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Browdy, Roger L.  
 ; REGISTRATION NUMBER: 25,618  
 ; REFERENCE/DOCKET NUMBER: TAKAKURA=1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 628-5197  
 ; TELEFAX: (202) 737-3528  
 ; INFORMATION FOR SEQ ID NO: 35:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 654 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; US-08-894-818B-35

Query Match 87.4%; Score 2997.5; DB 2; Length 654;  
 Best Local Similarity 88.3%; Pred. No. 8.5e-210;  
 Matches 582; Conservative 22; Mismatches 50; Indels 5; Gaps 1;

QY 1 MGKALKLIVLVGLVGSVAAAPPEKKEQVRNVEKNYGLTPGLFRKIQKLNPEEIS 60  
 DB 1 MGKALKLIVLVGLVGSVAAAPPEKKEQVRNVEKNYGLTPGLFRKIQKLNPEEIS 60  
 QY TVIVFENHREKEIAVRVLELMGAKRVYVYHIIIPAIADLVKVDLIVISGLTGSKAKLSGV 120  
 DB TVIVFENHREKEIAVRVLELMGAKRVYVYHIIIPAIADLVKVDLIVISGLTGSKAKLSGV 120  
 QY 121 RFIOEDYKTVSAELGLEDSAAQVMATYVWNLGYDGSGITIGIIDTGIDASHPDLOGKV 180  
 DB 121 RFIOEDYKTVSAELGLEDSAAQVMATYVWNLGYDGSGITIGIIDTGIDASHPDLOGKV 180

QY 181 ISWVDFVNRKSYPYDDHGHGHTVASIAAGTGAASNGKYKGMA PGAKLAIKYLGA DSGS 240  
 DB 181 ISWVDFVNRKSYPYDDHGHGHTVASIAAGTGAASNGKYKGMA PGAKLAIKYLGA DSGS 240  
 QY 241 ISTIIKGVAVDNKDCKGIGIKYINLSLSSQSGSDGTDSLQAVNNAMDGIYVCVAAGNS 300  
 DB 241 ISTIIKGVAVDNKDCKGIGIKYINLSLSSQSGSDGTDSLQAVNNAMDGIYVCVAAGNS 300  
 QY 301 GPNKTVGSRAASAKVITGAVDSDNNTASFSSRGPTADGRKPEVAVAPGVIIIPARASG 360  
 DB 301 GPNKTVGSRAASAKVITGAVDSDNNTASFSSRGPTADGRKPEVAVAPGVIIIPARASG 360  
 QY 361 TSMGTPINDYTTKASGSMATPHVSGVGLIIQAPHSWTPDKVTALJETADIAPKETA 420  
 DB 361 TSMGTPINDYTTKASGSMATPHVSGVGLIIQAPHSWTPDKVTALJETADIAPKETA 420  
 QY 421 DIAYGAGRVNYYKAIKYDDYAKLFTFGSVADKGSATHTFDVSGATFTVATLTLYWDTS SDI 480  
 DB 421 DIAYGAGRVNYYKAIKYDDYAKLFTFGSVADKGSATHTFDVSGATFTVATLTLYWDTS SDI 480  
 QY 481 DLXYDPNGNEVDYSTAYYGFEXGYNPTAGTWTVKVSYKGANVQVDVDSGLSQ 540  
 DB 481 DLXYDPNGNEVDYSTAYYGFEXGYNPTAGTWTVKVSYKGANVQVDVDSGLSQ 540  
 QY 541 SGGGNPNPNPNPTPTDTOTFTGTSVNDYMDTSDFTNANVSGATKINGDLTFDTSYND 600  
 DB 541 SGGGNPNPNPNPTPTDTOTFTGTSVNDYMDTSDFTNANVSGATKINGDLTFDTSYND 600  
 QY 541 -----FGSSPSPQEPETVYAKTFQGSDDHYVRSDFMTVNSGATKITGDVLFDSYND 595  
 DB 541 -----FGSSPSPQEPETVYAKTFQGSDDHYVRSDFMTVNSGATKITGDVLFDSYND 595  
 QY 601 LDLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFELVYASTYGMADYQLKAVVYVG 659  
 DB 596 LDLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFELVYASTYGMADYQLKAVVYVG 654

## RESULT 4

US-09-445-472-16  
 ; Sequence 16, Application US/09445472  
 ; Patent No. 6358726  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TAKAKURA, Hikaru  
 ; APPLICANT: MORISHITA, Mio  
 ; APPLICANT: SHIMOJO, Tomoko  
 ; APPLICANT: ASADA, Kiyozo  
 ; APPLICANT: KATO, Ikunoshin  
 ; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
 ; FILE REFERENCE: TAKAKURA=6  
 ; CURRENT APPLICATION NUMBER: US/09/445,472  
 ; CURRENT FILING DATE: 1999-12-06  
 ; PRIOR APPLICATION NUMBER: 151969/1997  
 ; PRIOR FILING DATE: 1997-06-10  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 16  
 ; LENGTH: 654  
 ; TYPE: PRT  
 ; ORGANISM: Pyrococcus furiosus  
 ; US-09-445-472-16

Query Match 87.4%; Score 2997.5; DB 2; Length 654;  
 Best Local Similarity 88.3%; Pred. No. 8.5e-210;  
 Matches 582; Conservative 22; Mismatches 50; Indels 5; Gaps 1;

QY 1 MGKALKLIVLVGLVGSVAAAPPEKKEQVRNVEKNYGLTPGLFRKIQKLNPEEIS 60  
 DB 1 MGKALKLIVLVGLVGSVAAAPPEKKEQVRNVEKNYGLTPGLFRKIQKLNPEEIS 60  
 QY TVIVFENHREKEIAVRVLELMGAKRVYVYHIIIPAIADLVKVDLIVISGLTGSKAKLSGV 120  
 DB TVIVFENHREKEIAVRVLELMGAKRVYVYHIIIPAIADLVKVDLIVISGLTGSKAKLSGV 120  
 QY 121 RFIOEDYKTVSAELGLEDSAAQVMATYVWNLGYDGSGITIGIIDTGIDASHPDLOGKV 180  
 DB 121 RFIOEDYKTVSAELGLEDSAAQVMATYVWNLGYDGSGITIGIIDTGIDASHPDLOGKV 180

```

QY 181 IGMVDFVNGRSYPPDDHGHGTHVASIAAGTGAASNGKYGMAFGAKLAGIKVLGADGSGS 240
DB 181 IGMVDFVNGRSYPPDDHGHGTHVASIAAGTGAASNGKYGMAFGAKLAGIKVLGADGSGS 240
QY 241 ISTIIKGVEMAVDNKDKYIGIKVINLSLGSQSSDGTDSLGAQVNNAMDAIIVVCAAGNS 300
DB 241 ISTIIKGVEMAVDNKDKYIGIKVINLSLGSQSSDGTDSLGAQVNNAMDAIIVVCAAGNS 300
QY 301 GPNNTYVSPAAASKVITVGAVDSDNDIASFSSRGPTADGRLKEPVVAPGVDIAPRASG 360
DB 301 GPNNTYVSPAAASKVITVGAVDSDNDIASFSSRGPTADGRLKEPVVAPGVDIAPRASG 360
QY 361 TSMGTPINDYTTKASGTSMATPHVSGVAGLIIQAPSWTPDKVKTALLETADIIVAPKEIA 420
DB 361 TSMGTPINDYTTKASGTSMATPHVSGVAGLIIQAPSWTPDKVKTALLETADIIVAPKEIA 420
QY 421 DIAYGAGRVNAYKAIKYDDYAKLTFTGSAVDKGSATHTEDVSGATPTVATLTYMDTGSSDI 480
DB 421 DIAYGAGRVNAYKAIKYDDYAKLTFTGSAVDKGSATHTEDVSGATPTVATLTYMDTGSSDI 480
QY 481 DLTYLDPNGNEVDYSYTAAYGFEKGYGNPTAGTWTVKVSYKGAANYQVDVSDGSLSQ 540
DB 481 DLTYLDPNGNEVDYSYTAAYGFEKGYGNPTAGTWTVKVSYKGAANYQVDVSDGSLSQ 540
QY 541 SGGGNPNPNPNPTPTDTQTFTGSAVDYMDTSDFTMNVNSGATKITGDLTFPTSYND 600
DB 541 SGGGNPNPNPNPTPTDTQTFTGSAVDYMDTSDFTMNVNSGATKITGDLTFPTSYND 600
QY 601 LDLYLDPNGNLYVDRSTSSNSYEHVEYANPAGTWTFLVYASTYGMADYOLKAVVYVG 659
DB 596 LDLYLDPNGNLYVDRSTSSNSYEHVEYANPAGTWTFLVYASTYGMADYOLKAVVYVG 654

```

## RESULT 5

```

US-10-090-624-16
; Sequence 16, Application US/10090624
; Patent No. 6783970
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-10-090-624-16

```

```

Query Match 87.4%; Score 2997.5; DB 2; Length 654;
Best Local Similarity 88.3%; Pred. No. 8,5e-10;
Matches 582; Conservative 22; Mismatches 50; Indels 5; Gaps 1;

```

```

QY 1 MKGILKLLIVILVIGLVGSAVAAPKPKVQVNVNKNYGLTPGLFRKIQKLNPEEIS 60
DB 1 MKGILKLLIVILVIGLVGSAVAAPKPKVQVNVNKNYGLTPGLFRKIQKLNPEEIS 60
QY 61 TVIVFENHREKEIAVAVLEIMGAKVYVYHIIIPAIADLKVRDLVIVISGLTGKAKLSGV 120
DB 61 TVIVFENHREKEIAVAVLEIMGAKVYVYHIIIPAIADLKVRDLVIVISGLTGKAKLSGV 120
QY 121 RFIOEDYKTVSALESLDESAAQVMATYVMNLGYGSGITIGIIDTGIDASHPDLQGRV 180

```

```

DB 121 RFIOEDYKTVSALESLDESAAQVMATYVMNLGYGSGITIGIIDTGIDASHPDLQGRV 180
QY 181 IGMVDFVNGRSYPPDDHGHGTHVASIAAGTGAASNGKYGMAFGAKLAGIKVLGADGSGS 240
DB 181 IGMVDFVNGRSYPPDDHGHGTHVASIAAGTGAASNGKYGMAFGAKLAGIKVLGADGSGS 240
QY 241 ISTIIKGVEMAVDNKDKYIGIKVINLSLGSQSSDGTDSLGAQVNNAMDAIIVVCAAGNS 300
DB 241 ISTIIKGVEMAVDNKDKYIGIKVINLSLGSQSSDGTDSLGAQVNNAMDAIIVVCAAGNS 300
QY 301 GPNNTYVSPAAASKVITVGAVDSDNDIASFSSRGPTADGRLKEPVVAPGVDIAPRASG 360
DB 301 GPNNTYVSPAAASKVITVGAVDSDNDIASFSSRGPTADGRLKEPVVAPGVDIAPRASG 360
QY 361 TSMGTPINDYTTKASGTSMATPHVSGVAGLIIQAPSWTPDKVKTALLETADIIVAPKEIA 420
DB 361 TSMGTPINDYTTKASGTSMATPHVSGVAGLIIQAPSWTPDKVKTALLETADIIVAPKEIA 420
QY 421 DIAYGAGRVNAYKAIKYDDYAKLTFTGSAVDKGSATHTEDVSGATPTVATLTYMDTGSSDI 480
DB 421 DIAYGAGRVNAYKAIKYDDYAKLTFTGSAVDKGSATHTEDVSGATPTVATLTYMDTGSSDI 480
QY 481 DLTYLDPNGNEVDYSYTAAYGFEKGYGNPTAGTWTVKVSYKGAANYQVDVSDGSLSQ 540
DB 481 DLTYLDPNGNEVDYSYTAAYGFEKGYGNPTAGTWTVKVSYKGAANYQVDVSDGSLSQ 540
QY 541 SGGGNPNPNPNPTPTDTQTFTGSAVDYMDTSDFTMNVNSGATKITGDLTFPTSYND 600
DB 541 SGGGNPNPNPNPTPTDTQTFTGSAVDYMDTSDFTMNVNSGATKITGDLTFPTSYND 600
QY 601 LDLYLDPNGNLYVDRSTSSNSYEHVEYANPAGTWTFLVYASTYGMADYOLKAVVYVG 659
DB 596 LDLYLDPNGNLYVDRSTSSNSYEHVEYANPAGTWTFLVYASTYGMADYOLKAVVYVG 654

```

## RESULT 6

```

US-09-841-553-35
; Sequence 35, Application US/09841553
; Patent No. 6849441
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masamori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,553
; FILING DATE: 24-Apr-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/894,818
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618

```

REFERENCE/DOCKET NUMBER: TAKAKURA=1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 654 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 35:  
US-09-841-553-35

Query Match 87.4%; Score 2997.5; DB 2; Length 654;  
Best Local Similarity 88.3%; Pred. No. 8.5e-210;  
Matches 582; Conservative 22; Mismatches 50; Indels 5; Gaps 1;

QY 1 MGKLKLLIIVILVLGLVGSVAAAPKPKYQVNNVEKNYGLTPGLFRKIQKLNPEEIS 60  
DB 1 MGKLKLLIIVILVLGLVGSVAAAPKPKYQVNNVEKNYGLTPGLFRKIQKLNPEEIS 60  
QY 61 TVIVFENHREKEIAVAVLELMGAKVRYVYHIIPAIADLKVRDLVLSGLTGKAKLSGV 120  
DB 61 TVIVFENHREKEIAVAVLELMGAKVRYVYHIIPAIADLKVRDLVLSGLTGKAKLSGV 120  
QY 121 RPIQEDYKTVSALEGLDESAAQVMAATVYMNIGYSGSGITIGIIDTGIDASHPDLOGRV 180  
DB 121 RPIQEDYKTVSALEGLDESAAQVMAATVYMNIGYSGSGITIGIIDTGIDASHPDLOGRV 180  
QY 181 IGVAVDPVNGRSYRYPDDHGHGTHVASIAAGTGAASNGKYGKMAPGAKLAGIKVLGADGSGS 240  
DB 181 IGVAVDPVNGRSYRYPDDHGHGTHVASIAAGTGAASNGKYGKMAPGAKLAGIKVLGADGSGS 240  
QY 241 ISTIIGVEMAVNDKDKYGIKINLSLGSQSSDGTDSLQAVNNMADAGIVCVAAGNS 300  
DB 241 ISTIIGVEMAVNDKDKYGIKINLSLGSQSSDGTDSLQAVNNMADAGIVCVAAGNS 300  
QY 301 GPKTYITIGSPAAASKYITVGAADVKKDYITSSSRGPTADRLKPEVVAPEMVIARASG 360  
DB 301 GPKTYITIGSPAAASKYITVGAADVKKDYITSSSRGPTADRLKPEVVAPEMVIARASG 360  
QY 361 TSMGPTINDYTTASGSMATPHVSGVGLILQAHPSWPDCKYKTALEIADIVAPKEIA 420  
DB 361 TSMGPTINDYTTASGSMATPHVSGVGLILQAHPSWPDCKYKTALEIADIVAPKEIA 420  
QY 421 DIAVGAQVNVYKAIKYDYAKLTFTGSVADKGSATHTPDVSGATFVTATLYMDTSSDI 480  
DB 421 DIAVGAQVNVYKAIKYDYAKLTFTGSVADKGSATHTPDVSGATFVTATLYMDTSSDI 480  
QY 481 DLVLYDPNGNEVDYSTYAYGFEKVGYNFTAGTWTVKVSYGGAANYOVVVSDGSLSQ 540  
DB 481 DLVLYDPNGNEVDYSTYAYGFEKVGYNFTAGTWTVKVSYGGAANYOVVVSDGSLSQ 540  
QY 541 SGGGNPNPNPNPPTPTTQTGTGSVNDWDTSDPTMNVNSGARKITGDLTFEDISYND 600  
DB 541 SGGGNPNPNPNPPTPTTQTGTGSVNDWDTSDPTMNVNSGARKITGDLTFEDISYND 600  
QY 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAGTWTFLVYAYSTYGADYQLKAVVYGG 659  
DB 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAGTWTFLVYAYSTYGADYQLKAVVYGG 659

APPLICANT: TSUNASAWA, Susumu  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Broadway and Neimark  
STREET: 419 Seventh Street N.W., Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/894,818B  
FILING DATE: 20-MAY-1998  
CLASSIFICATION: 435  
Prior Application DATA:  
APPLICATION NUMBER: PCT/JP96/03253  
FILING DATE: 07-NOV-1996  
Prior Application DATA:  
APPLICATION NUMBER: JP 323285/1995  
FILING DATE: 12-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TAKAKURA=1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 659 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-894-818B-1

Query Match 85.0%; Score 2914; DB 2; Length 659;  
Best Local Similarity 83.7%; Pred. No. 1e-203;  
Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps 7;

QY 1 MGKLKLLIIVILVLGLVGSVAAAPKPKYQVNN--VEKNYGLTPGLFRKIQKLNPE 57  
DB 1 MGKLKLLIIVILVLGLVGLAGTALAAPVKPY--VRNNAVQOKNYGLTPGLFRKIQKLNPE 58  
QY 58 EISTVIVFENHREKEIAVAVLELMGAKVRYVYHIIPAIADLKVRDLVLSGL--TG--G 113  
DB 58 EISTVIVFENHREKEIAVAVLELMGAKVRYVYHIIPAIADLKVRDLVLSGL--TG--G 113  
QY 114 KKLSGVRIQEDYKTVSALEGLDESAAQVMAATVYMNIGYSGSGITIGIIDTGIDIS 172  
DB 114 KKLSGVRIQEDYKTVSALEGLDESAAQVMAATVYMNIGYSGSGITIGIIDTGIDIS 172  
QY 173 HPDLQKVLGVWVDFVNGRSYRYPDDHGHGTHVASIAAGTGAASNGKYGKMAPGAKLAGIRV 232  
DB 173 HPDLQKVLGVWVDFVNGRSYRYPDDHGHGTHVASIAAGTGAASNGKYGKMAPGAKLAGIRV 232  
QY 233 LGADSGSISTIIKGVEMAVNDKDKYGIKINLSLGSQSSDGTDSLQAVNNMADAGIV 292  
DB 233 LGADSGSISTIIKGVEMAVNDKDKYGIKINLSLGSQSSDGTDSLQAVNNMADAGIV 292  
QY 293 VCVAAAGSGPNTYTVGSPPAAASKYITVGAVDNDNINIASFSRGPADGRLKPEVVAAPGV 352  
DB 293 VCVAAAGSGPNTYTVGSPPAAASKYITVGAVDNDNINIASFSRGPADGRLKPEVVAAPGV 352  
QY 353 IIAPRASGTSWGTPIINDYTTKASGSMATPHVSGVGLILQAHPSWTPDKYKTALEIAD 412  
DB 353 IIAPRASGTSWGTPIINDYTTKASGSMATPHVSGVGLILQAHPSWTPDKYKTALEIAD 412

```
QY 413 IYAPKEIADIAYAGAGRVNYKAIKYDDYAKLFTFGSVADKGSATHTFPVSGATFTATLY 472
Db 413 IYAPKEIADIAYAGAGRVNYKAIKYDDYAKLFTFGSVADKGSATHTFPVSGATFTATLY 472
QY 473 WDTGSSDIDLXYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTVAKVSYKGAANYQYDV 532
Db 473 WDTGSSDIDLXYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTVAKVSYKGAANYQYDV 532
QY 533 VSDGSLSGSGGPNPNPNPNPTPTDQTFTGSVNDWDTSDPTTMVNSGATKTTGDL 592
Db 533 VSDGSLSGSGGPNPNPNPNPTPTDQTFTGSVNDWDTSDPTTMVNSGATKTTGDL 592
QY 593 TFDTSYNDLXYDPNGNLVDRSTSSNSYEHVEYANPAPGTMTFLVYASTYGMADYQL 652
Db 593 TFDTSYNDLXYDPNGNLVDRSTSSNSYEHVEYANPAPGTMTFLVYASTYGMADYQL 652
QY 653 KAVVYVG 659
Db 653 KAVVYVG 659
```

## RESULT 8

```
US-09-445-472-12
; Sequence 12, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunobu
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445.472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Thermococcus celer
US-09-445-472-12
```

Query Match 85.0%; Score 2914; DB 2; Length 659;

Best Local Similarity 83.7%; Pred. No. 1e-203;

Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps 7;

```
QY 1 MKGLKALILVILVGLVYGVAAPEKKVEQVRN---VEKNYGLTFPGLFKRIQKLNPN 57
Db 1 MKRLGAVVLAIVVGLAGTALAAPVFPV--VRNNAVOQKRYGLTFGLFKKVRNMNQ 58
QY 58 EISTYIVFENHREKEIAVRVLELMGAKRVYHHIIPALADLVRLDLVLSGL--TG--G 113
Db 58 EVDTVIMFGSYGDRDAVKVRLMGAQVKSYKIIIPAVAVKIKARDLLLAGMTDGYFG 118
QY 114 KAKLSGVRFIOEDYKVTYSAELGLDESAQVMAITYWN--LGYDGSGITIGIITDGIDAS 172
Db 114 NTRVSGIKFIOEDYKQVDDA-----TSVQIGADTYMNSLGYDGSVVVAIVDTGIDAN 173
QY 173 HPDLQGVIGVWDFVNGRSYYPYDDHGHGTHVASIAAGTGAASNGRYKGMAPGAKLAGIV 232
Db 174 HPDLKGVIGVWDAVNGRSTPYDDQGHGTHVAGIVAGTGSV--NSQYIGVAPGAKLVGVKV 232
QY 233 LGADGSGSISTIIKGVEMAVDNKDYGKIVINLSLSSGSSDGTDSLGAANNAMDAIV 292
Db 233 LGADGSGSVSTIIAGVDMVYQNKDKGIRVINLSLSSGSSDGTDSLGAANNAMDAIV 292
QY 293 VCVAAAGNSGNTYTVSSPAASRVITVGAVDSDNNTIASFSSRGPTADGRLKPEVAPGYD 352
Db 293 VCVAAAGNSGNTYTVSSPAASRVITVGAVDSDNNTIASFSSRGPTADGRLKPEVAPGYD 352
```

```
QY 353 IIAFRASGTMGPINDYTTKASGTMATPHVSGVALLIIQAHPSWTPDKVTALLIETAD 412
Db 353 IIAFRASGTMGPINDYTTKASGTMATPHVSGVALLIIQAHPSWTPDKVTALLIETAD 412
QY 413 IYAPKEIADIAYAGAGRVNYKAIKYDDYAKLFTFGSVADKGSATHTFPVSGATFTATLY 472
Db 413 IYAPKEIADIAYAGAGRVNYKAIKYDDYAKLFTFGSVADKGSATHTFPVSGATFTATLY 472
QY 473 WDTGSSDIDLXYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTVAKVSYKGAANYQYDV 532
Db 473 WDTGSSDIDLXYDPNGNEVDYSYTAAYGFEKVGYNPTAGTWTVAKVSYKGAANYQYDV 532
QY 533 VSDGSLSGSGGPNPNPNPNPTPTDQTFTGSVNDWDTSDPTTMVNSGATKTTGDL 592
Db 533 VSDGSLSGSGGPNPNPNPNPTPTDQTFTGSVNDWDTSDPTTMVNSGATKTTGDL 592
QY 593 TFDTSYNDLXYDPNGNLVDRSTSSNSYEHVEYANPAPGTMTFLVYASTYGMADYQL 652
Db 593 TFDTSYNDLXYDPNGNLVDRSTSSNSYEHVEYANPAPGTMTFLVYASTYGMADYQL 652
QY 653 KAVVYVG 659
Db 653 KAVVYVG 659
```

## RESULT 9

```
US-10-090-624-12
; Sequence 12, Application US/10090624
; Patent No. 6783970
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunobu
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090.624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445.472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Thermococcus celer
US-10-090-624-12
```

Query Match 85.0%; Score 2914; DB 2; Length 659;

Best Local Similarity 83.7%; Pred. No. 1e-203;

Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps 7;

```
QY 1 MKGLKALILVILVGLVYGVAAPEKKVEQVRN---VEKNYGLTFPGLFKRIQKLNPN 57
Db 1 MKRLGAVVLAIVVGLAGTALAAPVFPV--VRNNAVOQKRYGLTFGLFKKVRNMNQ 58
QY 58 EISTYIVFENHREKEIAVRVLELMGAKRVYHHIIPALADLVRLDLVLSGL--TG--G 113
Db 58 EVDTVIMFGSYGDRDAVKVRLMGAQVKSYKIIIPAVAVKIKARDLLLAGMTDGYFG 118
QY 114 KAKLSGVRFIOEDYKVTYSAELGLDESAQVMAITYWN--LGYDGSGITIGIITDGIDAS 172
Db 114 NTRVSGIKFIOEDYKQVDDA-----TSVQIGADTYMNSLGYDGSVVVAIVDTGIDAN 173
QY 173 HPDLQGVIGVWDFVNGRSYYPYDDHGHGTHVASIAAGTGAASNGRYKGMAPGAKLAGIV 232
Db 174 HPDLKGVIGVWDAVNGRSTPYDDQGHGTHVAGIVAGTGSV--NSQYIGVAPGAKLVGVKV 232
QY 233 LGADGSGSISTIIKGVEMAVDNKDYGKIVINLSLSSGSSDGTDSLGAANNAMDAIV 292
Db 233 LGADGSGSVSTIIAGVDMVYQNKDKGIRVINLSLSSGSSDGTDSLGAANNAMDAIV 292
```







Db 416 QPEPTVADAKTFQXSDHYHYDRSDTFMTVNSGATKITGDLVFDTSYHDDLXYLDPNOKL 475  
QY 613 VDRSTSSNGEYEHYANPAPGTWTFVYASTYGADYOLKAVYYG 659  
Db 476 VDRSESPNSYEYEHYLTLPAGTWFLVYAYTYTGMAYYELTAKVYYG 522

## RESULT 13

US-10-090-624-4  
; Sequence 4, Application US/10090624  
; Patent No. 6783970  
; GENERAL INFORMATION:  
; APPLICANT: TAKAKURA, Hikaru  
; APPLICANT: MORISHITA, Mio  
; APPLICANT: SHIMOTO, Tomoko  
; APPLICANT: ASADA, Kiyozo  
; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
; FILE REFERENCE: TAKAKURA=6  
; CURRENT APPLICATION NUMBER: US/10/090,624  
; PRIOR FILING DATE: 2002-03-06  
; PRIOR APPLICATION NUMBER: 09/445,472  
; PRIOR FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: 151969/1997  
; PRIOR FILING DATE: 1997-06-10  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 4  
; LENGTH: 522  
; TYPE: PRT  
; ORGANISM: Pyrococcus furiosus  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (428)..(428)  
; OTHER INFORMATION: Xaa at position 428 is Gly or Val.  
US-10-090-624-4

Query Match 68.5%; Score 2349.5; DB 2; Length 522;  
Best Local Similarity 85.2%; Pred. No. 9.1e-163;  
Matches 449; Conservative 22; Mismatches 51; Indels 5; Gaps 1;

QY 133 AELEGDESAOVMATYVNNLGYDGSGITITGIDTGDASHPDLQGVIGWDFVNGRSY 192  
Db 1 AELEGDESAOVMATYVNNLGYDGSGITITGIDTGDASHPDLQGVIGWDFVNGRSY 60  
QY 193 PYDDHGHGTHVASTAGTGAASNGKTKGMAPAKLAGIKVILGADSGSISTIIKGVEMAV 252  
Db 61 PYDDHGHGTHVASTAGTGAASNGKTKGMAPAKLAGIKVILGADSGSISTIIKGVEMAV 120  
QY 253 DNKDKYGIKVINISLSSQSSDGTDSLQAVNNAMDAIGVCAAGNSGPNITYVGS PAA 312  
Db 121 DNKDKYGIKVINISLSSQSSDGTDSLQAVNNAMDAIGVCAAGNSGPNITYVGS PAA 180  
QY 313 ASKVITVGAVDSDNLTASFSRGPADGRKPEVNA PGVDIIPRASGSMGPINDYYT 372  
Db 181 ASKVITVGAVDKDYVTSFSRGPADGRKPEVNA PGVMWIIAARASGISMGPINDYYT 240  
QY 373 KASGTSMATPHVSGVALITLQAPHSWTPDKVKTALLETADIIVAPKEIADIAYAGRVNY 432  
Db 241 AAGTSMATPHVAGIALILQAPHSWTPDKVKTALLETADIIVAPKEIADIAYAGRVNY 300  
QY 433 KAIKYDDYAKLFTTGSVADKGSATHTFDVSGATFTATLTWDTGSSDIDLXYLDPNNGEV 492  
Db 301 KAINYNVAKLFTVGVANKGSQTHQFVLSGASFVATLTWDMANSDLDLXYLDPNNGEV 360  
QY 493 DVSATYAYGPEKGYNNPTAGTWTVYVSYKGAANOVDVSDSLSOSGSGPNPNPNP 552  
Db 361 DVSATYAYGPEKGYNNPTAGTWTVYVSYKGAANOVDVSDSLSO-----PGSFP 415  
QY 553 NPPTTDTOTFTGSVNDYWDYDSTFTMNVSAGTKITGDLTFDTSYNDLXYLDPNNGNL 612  
Db 416 QPEPTVADAKTFQXSDHYHYDRSDTFMTVNSGATKITGDLVFDTSYHDDLXYLDPNOKL 475

QY 613 VDRSTSSNGEYEHYANPAPGTWTFVYASTYGADYOLKAVYYG 659  
Db 476 VDRSESPNSYEYEHYLTLPAGTWFLVYAYTYTGMAYYELTAKVYYG 522

## RESULT 14

US-09-841-553-3  
; Sequence 3, Application US/09841553  
; Patent No. 6849441  
; GENERAL INFORMATION:  
; APPLICANT: TAKAKURA, Hikaru  
; APPLICANT: MORISHITA, Mio  
; APPLICANT: YAMAMOTO, Katsuhiko  
; APPLICANT: MURTA, Masanori  
; APPLICANT: ASADA, Kiyozo  
; APPLICANT: TSUNASAWA, Susumu  
; KATO, Ikunoshin  
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Broadway and Neimark  
; STREET: 419 Seventh Street N.W., Ste. 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: United States of America  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/841,553  
; FILING DATE: 24-Apr-2001  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/894,818  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: JP 323285/1995  
; FILING DATE: 12-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Broadway, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: TAKAKURA=1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 628-5197  
; TELEFAX: (202) 737-3528  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 522 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; OTHER INFORMATION:  
; Xaa at position 428 is Gly or Val.  
US-09-841-553-3

Query Match 68.5%; Score 2349.5; DB 2; Length 522;  
Best Local Similarity 85.2%; Pred. No. 9.1e-163;  
Matches 449; Conservative 22; Mismatches 51; Indels 5; Gaps 1;

QY 133 AELEGDESAOVMATYVNNLGYDGSGITITGIDTGDASHPDLQGVIGWDFVNGRSY 192  
Db 1 AELEGDESAOVMATYVNNLGYDGSGITITGIDTGDASHPDLQGVIGWDFVNGRSY 60  
QY 193 PYDDHGHGTHVASTAGTGAASNGKTKGMAPAKLAGIKVILGADSGSISTIIKGVEMAV 252  
Db 61 PYDDHGHGTHVASTAGTGAASNGKTKGMAPAKLAGIKVILGADSGSISTIIKGVEMAV 120  
QY 253 DNKDKYGIKVINISLSSQSSDGTDSLQAVNNAMDAIGVCAAGNSGPNITYVGS PAA 312

```

Db      121  DNKDYGIKVINLSIGSSQSDGTDALSOAVNAAMDAGLVVVAAGNSGPKYITIGSPAA 180
QY      313  ASKVIITVGAVDNSNNINISFSSRGPADGRLEKEVVAAPVDIIARASGTSMTINNYTT 372
Db      181  ASKVIITVGAVDKVIYITISFSSRGPADGRLEKEVVAAPVDIIARASGTSMTINNYTT 240
QY      373  KASGTSMTATPHVSGVAGLIIQAHPSMTPEKVKTALEIADIVAKELIADIVAGRVNVY 432
Db      241  AAPGTSMTATPHVAGIAALLIQAHPSMTPEKVKTALEIADIVAKELIADIVAGRVNVY 300
QY      433  KAIKVDYAKLTFTGVSADKSAATHTPEVSGATVVTATLTWDGSSPIDLYLDPNGNEV 492
Db      301  KAINVDYAKLVFTGYVANKSGQTHQFVISGASFVATLTWDMANSDLDLYLDPNGNOV 360
QY      493  DYSTAYYGFEEKVGYNPTAGTWTVKVSYKGAANYOVVSDGSLSSQSGGNPNPNP 552
Db      361  DYSTAYYGFEEKVGYNPTAGTWTIKVSYSGSANYOVVSDGSLSSQSGGNPNPNP 415
QY      553  NPFTTDTQFTGVSNDYMDTSDFTTMNVNSGATKLTGDLTFTDSYNDLTLVYDPNGNL 612
Db      416  QPEPTVDAKTFQXSDHYRSDFTMTVNSGATKLTGDLTFTDSYNDLTLVYDPNGNL 475
QY      613  VDRSTSGSYEHVYANPAPGTWELVYASTYGMADYOLKAVVYVG 659
Db      476  VDRSTSGSYEHVYANPAPGTWELVYASTYGMADYOLKAVVYVG 522

```

```

RESULT 15
US-09-445-472-1
/ Sequence 1, Application US/09445472
/ Patent No. 6358726
/ GENERAL INFORMATION:
/ APPLICANT: TAKAKURA, Hikaru
/ APPLICANT: MORISHITA, Mio
/ APPLICANT: SHIMOJO, Tomoko
/ APPLICANT: ASADA, Kiyozo
/ APPLICANT: KATO, Ikunoshin
/ TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
/ FILE REFERENCE: TAKAKURA=6
/ CURRENT APPLICATION NUMBER: US/09/445,472
/ CURRENT FILING DATE: 1999-12-06
/ PRIOR APPLICATION NUMBER: 151969/1997
/ PRIOR FILING DATE: 1997-06-10
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 1
/ LENGTH: 412
/ TYPE: PRF
/ ORGANISM: Pyrococcus furiosus
US-09-445-472-1

```

```

Query Match      56.0%; Score 1918; DB 2; Length 412;
Best Local Similarity 89.5%; Pred. No. 1.7e-131;
Matches 367; Conservative 15; Mismatches 28; Indels 0; Gaps 0;

```

```

QY      133  AELIEGIDESAQVMAATYVWNLGYDGSGITIGITDGTIDASHPLDQKVTIGWVDFVNGRSY 192
Db      1  AELIEGIDESAQVMAATYVWNLGYDGSGITIGITDGTIDASHPLDQKVTIGWVDFVNGRSY 60
QY      193  PYDDHGHGTHVASIAAGTGAASNGKYGAPAGKLAGIKVILGADSGSISTIIKGVEMAV 252
Db      61  PYDDHGHGTHVASIAAGTGAASNGKYGAPAGKLAGIKVILGADSGSISTIIKGVEMAV 120
QY      253  DNKDKYGIKVINLSIGSSQSDGTDLSQAVNNAMDAIIVCAAGNSGPNPTVYVSPAA 312
Db      121  DNKDKYGIKVINLSIGSSQSDGTDLSQAVNNAMDAIIVCAAGNSGPNPTVYVSPAA 180
QY      313  ASKVIITVGAVDNSNNINISFSSRGPADGRLEKEVVAAPVDIIARASGTSMTINNYTT 372
Db      181  ASKVIITVGAVDKVIYITISFSSRGPADGRLEKEVVAAPVDIIARASGTSMTINNYTT 240
QY      373  KASGTSMTATPHVSGVAGLIIQAHPSMTPEKVKTALEIADIVAKELIADIVAGRVNVY 432

```

```

Db      241  AAPGTSMTATPHVAGIAALLIQAHPSMTPEKVKTALEIADIVAKELIADIVAGRVNVY 300
QY      433  KAIKVDYAKLTFTGVSADKSAATHTPEVSGATVVTATLTWDGSSPIDLYLDPNGNEV 492
Db      301  KAINVDYAKLVFTGYVANKSGQTHQFVISGASFVATLTWDMANSDLDLYLDPNGNOV 360
QY      493  DYSTAYYGFEEKVGYNPTAGTWTVKVSYKGAANYOVVSDGSLSSQSG 542
Db      361  DYSTAYYGFEEKVGYNPTAGTWTIKVSYSGSANYOVVSDGSLSSQSG 410

```

Search completed: January 6, 2007, 22:33:38  
Job time : 38 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2007, 22:32:37 ; Search time 123.5 Seconds  
(without alignments)  
2471.729 Million cell updates/sec

Title: US-10-800-684-1  
Perfect score: 3437  
Sequence: 1 MRLGAVLAVLVGLACT.....YASTYGMADYQLKAVVYG 659

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_Main:  
1: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US07\_PUBSCOMB.pep.\*  
2: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US08\_PUBSCOMB.pep.\*  
3: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US09\_PUBSCOMB.pep.\*  
4: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US10A\_PUBSCOMB.pep.\*  
5: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US10B\_PUBSCOMB.pep.\*  
6: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US11\_PUBSCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score  | Query Match | Length | DB | ID                  | Description       |
|------------|--------|-------------|--------|----|---------------------|-------------------|
| 1          | 3437   | 100.0       | 659    | 3  | US-09-841-553-1     | Sequence 1, Appl  |
| 2          | 3437   | 100.0       | 659    | 4  | US-10-090-624-12    | Sequence 12, Appl |
| 3          | 3437   | 100.0       | 659    | 5  | US-10-800-684-1     | Sequence 1, Appl  |
| 4          | 3437   | 100.0       | 659    | 5  | US-10-888-588-12    | Sequence 12, Appl |
| 5          | 2914   | 84.8        | 659    | 5  | US-09-841-553-5     | Sequence 5, Appl  |
| 6          | 2914   | 84.8        | 659    | 5  | US-10-800-684-5     | Sequence 5, Appl  |
| 7          | 2483.5 | 72.3        | 654    | 3  | US-09-841-553-35    | Sequence 35, Appl |
| 8          | 2483.5 | 72.3        | 654    | 4  | US-10-090-624-16    | Sequence 16, Appl |
| 9          | 2483.5 | 72.3        | 654    | 5  | US-10-800-684-35    | Sequence 35, Appl |
| 10         | 2483.5 | 72.3        | 654    | 5  | US-10-888-588-16    | Sequence 16, Appl |
| 11         | 2138.5 | 62.2        | 522    | 3  | US-09-841-553-3     | Sequence 3, Appl  |
| 12         | 2138.5 | 62.2        | 522    | 4  | US-10-090-624-4     | Sequence 4, Appl  |
| 13         | 2138.5 | 62.2        | 522    | 5  | US-10-800-684-3     | Sequence 3, Appl  |
| 14         | 2138.5 | 62.2        | 522    | 5  | US-10-888-588-4     | Sequence 4, Appl  |
| 15         | 1707   | 49.7        | 412    | 4  | US-10-090-624-1     | Sequence 1, Appl  |
| 16         | 1707   | 49.7        | 412    | 5  | US-10-888-588-1     | Sequence 1, Appl  |
| 17         | 668.5  | 19.5        | 1139   | 4  | US-10-156-761-10856 | Sequence 10856, A |
| 18         | 657.5  | 19.1        | 1208   | 4  | US-10-156-761-13251 | Sequence 13251, A |
| 19         | 636    | 18.5        | 1079   | 4  | US-10-112-488-39    | Sequence 39, Appl |
| 20         | 636    | 18.5        | 1079   | 4  | US-10-673-860-8     | Sequence 8, Appl  |
| 21         | 632.5  | 18.4        | 1237   | 5  | US-10-314-657-4     | Sequence 4, Appl  |
| 22         | 632.5  | 18.4        | 1237   | 5  | US-10-473-193-4     | Sequence 4, Appl  |
| 23         | 564.5  | 16.4        | 379    | 3  | US-09-813-408-11    | Sequence 11, Appl |
| 24         | 563.5  | 16.4        | 379    | 6  | US-11-156-062-23    | Sequence 23, Appl |
| 25         | 560.5  | 16.3        | 379    | 6  | US-10-202-339-2     | Sequence 2, Appl  |
| 26         | 559.5  | 16.3        | 379    | 3  | US-09-920-118-14    | Sequence 14, Appl |
| 27         | 557.5  | 16.2        | 379    | 3  | US-09-813-408-10    | Sequence 10, Appl |

|    |       |      |      |   |                   |                   |
|----|-------|------|------|---|-------------------|-------------------|
| 28 | 557.5 | 16.2 | 379  | 4 | US-10-146-905A-10 | Sequence 10, Appl |
| 29 | 557.5 | 16.2 | 379  | 5 | US-10-503-706-1   | Sequence 1, Appl  |
| 30 | 557.5 | 16.2 | 380  | 4 | US-10-146-905A-12 | Sequence 12, Appl |
| 31 | 556.5 | 16.2 | 379  | 3 | US-09-813-408-12  | Sequence 12, Appl |
| 32 | 554.5 | 16.1 | 372  | 3 | US-09-813-408-9   | Sequence 9, Appl  |
| 33 | 551.5 | 16.0 | 378  | 3 | US-09-813-408-14  | Sequence 14, Appl |
| 34 | 550.5 | 16.0 | 379  | 3 | US-09-813-408-13  | Sequence 13, Appl |
| 35 | 547   | 15.9 | 382  | 3 | US-09-957-806A-10 | Sequence 10, Appl |
| 36 | 546   | 15.9 | 338  | 3 | US-09-813-408-7   | Sequence 7, Appl  |
| 37 | 538   | 15.7 | 310  | 3 | US-09-813-408-15  | Sequence 15, Appl |
| 38 | 535.5 | 15.6 | 1398 | 3 | US-09-841-553-8   | Sequence 8, Appl  |
| 39 | 535.5 | 15.6 | 1398 | 4 | US-10-090-624-6   | Sequence 6, Appl  |
| 40 | 535.5 | 15.6 | 1398 | 5 | US-10-800-684-8   | Sequence 8, Appl  |
| 41 | 534.5 | 15.6 | 1398 | 5 | US-10-888-588-6   | Sequence 6, Appl  |
| 42 | 534.5 | 15.6 | 382  | 4 | US-10-146-905A-11 | Sequence 11, Appl |
| 43 | 533.5 | 15.5 | 382  | 4 | US-10-090-624-31  | Sequence 31, Appl |
| 44 | 533.5 | 15.5 | 382  | 4 | US-10-104-693-2   | Sequence 2, Appl  |
| 45 | 533.5 | 15.5 | 382  | 4 | US-10-146-905A-8  | Sequence 8, Appl  |

## ALIGNMENTS

RESULT 1  
US-09-841-553-1  
Sequence 1, Application US/09841553  
Publication No. US20020086402A1  
GENERAL INFORMATION:  
APPLICANT: TAKAKURA, HiKaru  
MORISHITA, Mio  
YAMAMOTO, Katsuniko  
MIRTA, Masanori  
ASADA, Kiyozo  
TSUNASAKA, Susumu  
KATO, Ikumoshin  
TITLE OF INVENTION: HYPERHERMOSTABLE PROTEASE GENES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESSES:  
ADDRESS: Broadway and Neilmark  
STREET: 419 Seventh Street N.W., Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/841,553  
FILING DATE: 24-Apr-2001  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/894,818  
FILING DATE: <Unknown>  
APPLICATION NUMBER: JP 323285/1995  
FILING DATE: 12-Dec-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TAKAKURA=1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 659 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-841-553-1

*Handwritten notes:*  
"Dns of this appl. issued as US 6,849,444"  
"In re claims only DNA"  
"No Pbl-pat."

Query Match 100.0%; Score 3437; DB 3; Length 659;  
Best Local Similarity 100.0%; Pred. No. 9,3e-242;  
Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MKRLGAVVLAIVVGLLAGTALAAPVFPVYRNAVVOQKNGYGLTTPGLFKYQRMNNQEV 60
DB 1 MKRLGAVVLAIVVGLLAGTALAAPVFPVYRNAVVOQKNGYGLTTPGLFKYQRMNNQEV 60
QY 61 DTVMFSGYDGRDRAVKVRLMGAQVYKXIIIPAVAVKIKARDLLIAGMIDTGYGNT 120
DB 61 DTVMFSGYDGRDRAVKVRLMGAQVYKXIIIPAVAVKIKARDLLIAGMIDTGYGNT 120
QY 121 RVSGIRFIOEDYKVOVDATSVQIGADTVNNSLGVDGQVVAIVDTGIDANHPDLKXK 180
DB 121 RVSGIRFIOEDYKVOVDATSVQIGADTVNNSLGVDGQVVAIVDTGIDANHPDLKXK 180
QY 181 VIGMYDAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGS 240
DB 181 VIGMYDAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGS 240
QY 241 VSTIIAGVDMVYQKDKYKIRVINLSLSSQSSDGTDSLQAVNNAMDAGIIVCVAAAGNS 300
DB 241 VSTIIAGVDMVYQKDKYKIRVINLSLSSQSSDGTDSLQAVNNAMDAGIIVCVAAAGNS 300
QY 301 GPNTYTVGSPAASKVITVGAVDNDNIASFSSRGPTADGRLKPEVVAPEGVDIIAPRAGS 360
DB 301 GPNTYTVGSPAASKVITVGAVDNDNIASFSSRGPTADGRLKPEVVAPEGVDIIAPRAGS 360
QY 361 TSMGTPINDYTTKASGTSMAITPHVSGVALLIQAHPMTDPKVTALILETADIAPAEIA 420
DB 361 TSMGTPINDYTTKASGTSMAITPHVSGVALLIQAHPMTDPKVTALILETADIAPAEIA 420
QY 421 DIAVGAQVNVYKAIKYDDYAKLFTGSAVDKGSATHTPDVSGATFTVATLWDTGSSDI 480
DB 421 DIAVGAQVNVYKAIKYDDYAKLFTGSAVDKGSATHTPDVSGATFTVATLWDTGSSDI 480
QY 481 DLVLYDPNGNEVDYSTAYYGFEEKVGYNPTAGTMTVKVSYKGAANYQVDVSDSLQ 540
DB 481 DLVLYDPNGNEVDYSTAYYGFEEKVGYNPTAGTMTVKVSYKGAANYQVDVSDSLQ 540
QY 541 SGGGNPNPNPNPTPTTDTQFTGSVNDYWDSDTFTMNVNSGATKITGDLTFDTSYND 600
DB 541 SGGGNPNPNPNPTPTTDTQFTGSVNDYWDSDTFTMNVNSGATKITGDLTFDTSYND 600
QY 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTMTFLVYASTYGMADYQLKAVVYGG 659
DB 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTMTFLVYASTYGMADYQLKAVVYGG 659
```

## RESULT 2

```
US-10-090-624-12
; Sequence 12, Application US/10090624
; Publication No. US2002013235A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 659
; TYPE: PRT
```

ORGANISM: Thermococcus celer  
US-10-090-624-12

Query Match 100.0%; Score 3437; DB 4; Length 659;  
Best Local Similarity 100.0%; Pred. No. 9,3e-242;  
Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MKRLGAVVLAIVVGLLAGTALAAPVFPVYRNAVVOQKNGYGLTTPGLFKYQRMNNQEV 60
DB 1 MKRLGAVVLAIVVGLLAGTALAAPVFPVYRNAVVOQKNGYGLTTPGLFKYQRMNNQEV 60
QY 61 DTVMFSGYDGRDRAVKVRLMGAQVYKXIIIPAVAVKIKARDLLIAGMIDTGYGNT 120
DB 61 DTVMFSGYDGRDRAVKVRLMGAQVYKXIIIPAVAVKIKARDLLIAGMIDTGYGNT 120
QY 121 RVSGIRFIOEDYKVOVDATSVQIGADTVNNSLGVDGQVVAIVDTGIDANHPDLKXK 180
DB 121 RVSGIRFIOEDYKVOVDATSVQIGADTVNNSLGVDGQVVAIVDTGIDANHPDLKXK 180
QY 181 VIGMYDAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGS 240
DB 181 VIGMYDAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGS 240
QY 241 VSTIIAGVDMVYQKDKYKIRVINLSLSSQSSDGTDSLQAVNNAMDAGIIVCVAAAGNS 300
DB 241 VSTIIAGVDMVYQKDKYKIRVINLSLSSQSSDGTDSLQAVNNAMDAGIIVCVAAAGNS 300
QY 301 GPNTYTVGSPAASKVITVGAVDNDNIASFSSRGPTADGRLKPEVVAPEGVDIIAPRAGS 360
DB 301 GPNTYTVGSPAASKVITVGAVDNDNIASFSSRGPTADGRLKPEVVAPEGVDIIAPRAGS 360
QY 361 TSMGTPINDYTTKASGTSMAITPHVSGVALLIQAHPMTDPKVTALILETADIAPAEIA 420
DB 361 TSMGTPINDYTTKASGTSMAITPHVSGVALLIQAHPMTDPKVTALILETADIAPAEIA 420
QY 421 DIAVGAQVNVYKAIKYDDYAKLFTGSAVDKGSATHTPDVSGATFTVATLWDTGSSDI 480
DB 421 DIAVGAQVNVYKAIKYDDYAKLFTGSAVDKGSATHTPDVSGATFTVATLWDTGSSDI 480
QY 481 DLVLYDPNGNEVDYSTAYYGFEEKVGYNPTAGTMTVKVSYKGAANYQVDVSDSLQ 540
DB 481 DLVLYDPNGNEVDYSTAYYGFEEKVGYNPTAGTMTVKVSYKGAANYQVDVSDSLQ 540
QY 541 SGGGNPNPNPNPTPTTDTQFTGSVNDYWDSDTFTMNVNSGATKITGDLTFDTSYND 600
DB 541 SGGGNPNPNPNPTPTTDTQFTGSVNDYWDSDTFTMNVNSGATKITGDLTFDTSYND 600
QY 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTMTFLVYASTYGMADYQLKAVVYGG 659
DB 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTMTFLVYASTYGMADYQLKAVVYGG 659
```

## RESULT 3

```
US-10-800-684-1
; Sequence 1, Application US/10800684
; Publication No. US20050014221A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: ULTRATHERMOSTABLE PROTEASE GENES
; FILE REFERENCE: TAKAKURA=1A
; CURRENT APPLICATION NUMBER: US/10/800,684
; PRIOR APPLICATION NUMBER: 2004-03-16
; PRIOR APPLICATION NUMBER: US/09/841,553
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 08/894,818
; PRIOR FILING DATE: 1997-08-29
; PRIOR APPLICATION NUMBER: JP32385/1995
```

2

Did not claim Seq ID 12, 3 & 4.  
which has not been taken  
into DH file

PRIOR FILING DATE: 1995-12-12  
PRIOR APPLICATION NUMBER: JP96/03253  
PRIOR FILING DATE: 1996-11-07  
NUMBER OF SEQ ID NOS: 45  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 1  
LENGTH: 659  
TYPE: PR  
ORGANISM: Thermococcus celer  
US-10-800-684-1

*Handwritten:* This is not a claim. It is a sequence. No. 1. Not a claim.

Query Match 100.0%; Score 3437; DB 5; Length 659;  
Best Local Similarity 100.0%; Pred. No. 9.3e-242;  
Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRLGAVLALVYVGLAGTALAAPYKPVVNNNAVOOKYGLTTPGLFKVQRMNNQEV 60  
DB 1 MKRLGAVLALVYVGLAGTALAAPYKPVVNNNAVOOKYGLTTPGLFKVQRMNNQEV 60  
QY 61 DTVMGSGYGDRAVAVKRLMGAQVYSYKIIIPAAVAKIKARDLLIAGMIDTGYFGNT 120  
DB 61 DTVMGSGYGDRAVAVKRLMGAQVYSYKIIIPAAVAKIKARDLLIAGMIDTGYFGNT 120  
QY 121 RVSGIRFIEDYKQVVDATSVQIGADTYMNSLGVDGSSVVAIVDTGIDANHPDLKKG 180  
DB 121 RVSGIRFIEDYKQVVDATSVQIGADTYMNSLGVDGSSVVAIVDTGIDANHPDLKKG 180  
QY 181 VIGWYAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGKVLGADGSGS 240  
DB 181 VIGWYAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGKVLGADGSGS 240  
QY 241 VSTIIAGVDMVQNKDKYGRVNLISGSSQSSDGTDSLQAVNNMADAGIIVCVAAGNS 300  
DB 241 VSTIIAGVDMVQNKDKYGRVNLISGSSQSSDGTDSLQAVNNMADAGIIVCVAAGNS 300  
QY 301 GPRTYVGSPPAASAKYITVGAVDSDNINIASFSSRGPTADGRLEPVVAPGVDIIPRAGS 360  
DB 301 GPRTYVGSPPAASAKYITVGAVDSDNINIASFSSRGPTADGRLEPVVAPGVDIIPRAGS 360  
QY 361 TSMGTPINDYTKASGTSMAATPHVSGVGLIIQAHPSMTDPDKYKIALIETADIAPKEIA 420  
DB 361 TSMGTPINDYTKASGTSMAATPHVSGVGLIIQAHPSMTDPDKYKIALIETADIAPKEIA 420  
QY 421 DIAVGAARNVVAIKYDDYAKLTFGSAVDKGSATHTFVSGATFVTATLYMDTSSDI 480  
DB 421 DIAVGAARNVVAIKYDDYAKLTFGSAVDKGSATHTFVSGATFVTATLYMDTSSDI 480  
QY 481 DLYLDPNGNEVDYSTAYYGFEEKVGYNPTAGTWKVVSYGGAANYQVDVSDGSLSQ 540  
DB 481 DLYLDPNGNEVDYSTAYYGFEEKVGYNPTAGTWKVVSYGGAANYQVDVSDGSLSQ 540  
QY 541 SGGGNPNPNPNPPTTDTOTFTGSVNDYWDTSDFTMNVNSGATKIGDLTFDTSYND 600  
DB 541 SGGGNPNPNPNPPTTDTOTFTGSVNDYWDTSDFTMNVNSGATKIGDLTFDTSYND 600  
QY 601 LDLYLDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVLYAVSTYGMADYQLAAVYYG 659  
DB 601 LDLYLDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVLYAVSTYGMADYQLAAVYYG 659

RESULT 4  
US-10-888-588-12

Sequence 12, Application US/10888588  
Publication No. US2005084934A1  
GENERAL INFORMATION:  
APPLICANT: TAKAKURA, Hikaru  
APPLICANT: MORISHITA, Mio  
APPLICANT: SHIMOJO, Tomoko  
APPLICANT: ASADA, Kiyozo  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
FILE REFERENCE: TAKAKURA-6  
CURRENT APPLICATION NUMBER: US/10/888,588

CURRENT FILING DATE: 2004-07-12  
PRIOR APPLICATION NUMBER: US/10/090,624  
PRIOR FILING DATE: 2002-03-06  
PRIOR APPLICATION NUMBER: 09/445,472  
PRIOR FILING DATE: 1999-12-06  
PRIOR APPLICATION NUMBER: 151969/1997  
PRIOR FILING DATE: 1997-06-10  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 12  
LENGTH: 659  
TYPE: PR  
ORGANISM: Thermococcus celer  
US-10-888-588-12

*Handwritten:* This is not a claim. It is a sequence. No. 1. Not a claim.

Query Match 100.0%; Score 3437; DB 5; Length 659;  
Best Local Similarity 100.0%; Pred. No. 9.3e-242;  
Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRLGAVLALVYVGLAGTALAAPYKPVVNNNAVOOKYGLTTPGLFKVQRMNNQEV 60  
DB 1 MKRLGAVLALVYVGLAGTALAAPYKPVVNNNAVOOKYGLTTPGLFKVQRMNNQEV 60  
QY 61 DTVMGSGYGDRAVAVKRLMGAQVYSYKIIIPAAVAKIKARDLLIAGMIDTGYFGNT 120  
DB 61 DTVMGSGYGDRAVAVKRLMGAQVYSYKIIIPAAVAKIKARDLLIAGMIDTGYFGNT 120  
QY 121 RVSGIRFIEDYKQVVDATSVQIGADTYMNSLGVDGSSVVAIVDTGIDANHPDLKKG 180  
DB 121 RVSGIRFIEDYKQVVDATSVQIGADTYMNSLGVDGSSVVAIVDTGIDANHPDLKKG 180  
QY 181 VIGWYAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGKVLGADGSGS 240  
DB 181 VIGWYAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGKVLGADGSGS 240  
QY 241 VSTIIAGVDMVQNKDKYGRVNLISGSSQSSDGTDSLQAVNNMADAGIIVCVAAGNS 300  
DB 241 VSTIIAGVDMVQNKDKYGRVNLISGSSQSSDGTDSLQAVNNMADAGIIVCVAAGNS 300  
QY 301 GPRTYVGSPPAASAKYITVGAVDSDNINIASFSSRGPTADGRLEPVVAPGVDIIPRAGS 360  
DB 301 GPRTYVGSPPAASAKYITVGAVDSDNINIASFSSRGPTADGRLEPVVAPGVDIIPRAGS 360  
QY 361 TSMGTPINDYTKASGTSMAATPHVSGVGLIIQAHPSMTDPDKYKIALIETADIAPKEIA 420  
DB 361 TSMGTPINDYTKASGTSMAATPHVSGVGLIIQAHPSMTDPDKYKIALIETADIAPKEIA 420  
QY 421 DIAVGAARNVVAIKYDDYAKLTFGSAVDKGSATHTFVSGATFVTATLYMDTSSDI 480  
DB 421 DIAVGAARNVVAIKYDDYAKLTFGSAVDKGSATHTFVSGATFVTATLYMDTSSDI 480  
QY 481 DLYLDPNGNEVDYSTAYYGFEEKVGYNPTAGTWKVVSYGGAANYQVDVSDGSLSQ 540  
DB 481 DLYLDPNGNEVDYSTAYYGFEEKVGYNPTAGTWKVVSYGGAANYQVDVSDGSLSQ 540  
QY 541 SGGGNPNPNPNPPTTDTOTFTGSVNDYWDTSDFTMNVNSGATKIGDLTFDTSYND 600  
DB 541 SGGGNPNPNPNPPTTDTOTFTGSVNDYWDTSDFTMNVNSGATKIGDLTFDTSYND 600  
QY 601 LDLYLDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVLYAVSTYGMADYQLAAVYYG 659  
DB 601 LDLYLDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFVLYAVSTYGMADYQLAAVYYG 659

RESULT 5  
US-09-841-553-5

Sequence 5, Application US/09841553  
Publication No. US20020086402A1  
GENERAL INFORMATION:  
APPLICANT: TAKAKURA, Hikaru  
APPLICANT: MORISHITA, Mio  
APPLICANT: YAMAMOTO, Katsuniko  
MITTA, Masaori

```

ASADA, Kiyozo
TSUNASAWA, Susumu
KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESS: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/841,553
FILING DATE: 24-Apr-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/894,818
FILING DATE: <Unknown>
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 5
SEQUENCE CHARACTERISTICS:
LENGTH: 659 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 5
US-09-841-553-5

Query Match      84.8%; Score 2914; DB 3; Length 659;
Best Local Similarity 83.7%; Pred. No. 1.3e-203;
Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps 7;

QY 1 MKRLGAVVLAIVLVLGTLAALAPVKPV--VRNNAVQOKNYGLITPGLFKKYQRMNMNQ 58
DB 1 MKGLKALILVLVLGLVGVSAAPBEKKVEQVRN---VEKNYGLITPGLFKKIQKLNPN 57
QY 59 EVDVIVMGSGYGDRAVAVKYLRLMGAVKYSYKIIPAVAVKIKARDLLIAGMIDTGYFG 118
DB 58 EISTVIVFENHREKEIAVRVLELMGAKRVYVHIIPALADLKVRDLVLSGL--TG--G 113
QY 119 NTRVSGIKFTIOEDYKVVQVDA-----TSVSGIAGDTVWNSLGYDGSVVVAIVDTGIDAN 173
DB 114 KAKLSGVRFIOEDYKVVSALEGLDESAAQVMAITYWN--LGIDGSGITIGIITGIDAS 172
QY 174 HPDLKGVITGVDAVNGRSTPYDDQGHGTHVAGIVAGTGSV-NSQYIGVAPGAKLVGVKV 232
DB 173 HPDLQGVITGVDAVNGRSTPYDDHGHGTHVASIAAGTGAASNGKYGMAFGAKLGITKV 232
QY 233 LGADGSGSVSTIIAGVDVWVQNDKYGIRVINSLSGSSGSDGTDLSQAVNNAMAGIV 292
DB 233 LGADGSGSVSTIIKGVMAVNDKQGIKVINLSLSSGSDGTDLSQAVNNAMAGIV 292
QY 293 VCVAAAGNSGNTYTVGSPPAASKVITVGAVDSDNDNTASFSSRGPRTADGRLKEPVAPVD 352
DB 293 VCVAAAGNSGNTYTVGSPPAASKVITVGAVDSDNDNTASFSSRGPRTADGRLKEPVAPVD 352
QY 353 IIPRASGTSMTGPIINDYTKASGTSMAIPHVSQVALLIQAHPSTPDKVTALITETAD 412
DB 353 IIPRASGTSMTGPIINDYTKASGTSMAIPHVSQVALLIQAHPSTPDKVTALITETAD 412

```

```

QY 413 IVAPKEIADTAYGAGRVNVTAKIKYDDYAKLTFTFGSVADKGSATHTFDVSGATFVATLLY 472
DB 413 IVAPKEIADTAYGAGRVNVTAKIKYDDYAKLTFTFGSVADKGSATHTFDVSGATFVATLLY 472
QY 473 WDTSDDIDLYLVDPNNGNEVDYSYTAAYGFEKGYVNPAGTMTVKKVSKGAANOVDV 532
DB 473 WDTSDDIDLYLVDPNNGNEVDYSYTAAYGFEKGYVNPAGTMTVKKVSKGAANOVDV 532
QY 533 VSDGSLSGSGGNPNPNPNPTPTDQTQFTGSVNDYWDTSPTFTMNVSAGATKITGDL 592
DB 533 VSDGSLSGSGGNPNPNPNPTPTDQTQFTGSVNDYWDTSPTFTMNVSAGATKITGDL 592
QY 593 TFPDSVNDLILYLDPNVGNLVDKSTSSNSYEHTEYANPAPGWTPLVYASTGADYQL 652
DB 593 TFPDSVNDLILYLDPNVGNLVDKSTSSNSYEHTEYANPAPGWTPLVYASTGADYQL 652
QY 653 KAVVYYG 659
DB 653 KAVVYYG 659

RESULT 6
US-10-800-684-5
Sequence 5, Application US/10800684
Publication No. US20050014221A1
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: ULTRATHERMOSTABLE PROTEASE GENES
FILE REFERENCE: TAKAKURA=1A
CURRENT APPLICATION NUMBER: US/10/800,684
CURRENT FILING DATE: 2004-03-16
PRIOR APPLICATION NUMBER: US-09/841,553
PRIOR FILING DATE: 2001-04-24
PRIOR APPLICATION NUMBER: 08/894,818
PRIOR FILING DATE: 1997-08-29
PRIOR APPLICATION NUMBER: JP32385/1995
PRIOR FILING DATE: 1995-12-12
PRIOR APPLICATION NUMBER: JP96/03253
PRIOR FILING DATE: 1996-11-07
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn version 3.2
SEQ ID NO 5
LENGTH: 659
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-10-800-684-5

Query Match      84.8%; Score 2914; DB 5; Length 659;
Best Local Similarity 83.7%; Pred. No. 1.3e-203;
Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps 7;

QY 1 MKRLGAVVLAIVLVLGTLAALAPVKPV--VRNNAVQOKNYGLITPGLFKKYQRMNMNQ 58
DB 1 MKGLKALILVLVLGLVGVSAAPBEKKVEQVRN---VEKNYGLITPGLFKKIQKLNPN 57
QY 59 EVDVIVMGSGYGDRAVAVKYLRLMGAVKYSYKIIPAVAVKIKARDLLIAGMIDTGYFG 118
DB 58 EISTVIVFENHREKEIAVRVLELMGAKRVYVHIIPALADLKVRDLVLSGL--TG--G 113
QY 119 NTRVSGIKFTIOEDYKVVQVDA-----TSVSGIAGDTVWNSLGYDGSVVVAIVDTGIDAN 173
DB 114 KAKLSGVRFIOEDYKVVSALEGLDESAAQVMAITYWN--LGIDGSGITIGIITGIDAS 172
QY 174 HPDLKGVITGVDAVNGRSTPYDDQGHGTHVAGIVAGTGSV-NSQYIGVAPGAKLVGVKV 232

```



```

Db 173 HPDLQKVLGWVDFVNGRSYPYDDHGHGTHVASIAAGTGAASNGKKGAPAKLAGIKV 232
QY 233 LGADSGSVSTIIAGVWVQNKDKYKIRYINLSLSSOSSDCTDLSQAVNNAMAGIV 292
Db 233 LGADSGSVSTIIAGVWVQNKDKYKIRYINLSLSSOSSDCTDLSQAVNNAMAGIV 292
QY 293 VCVAAAGNSGPNYTVGSPAAASKVITVGAVDSDNINASFSSRGPTADGRLKEPVAPGV 352
Db 293 VCVAAAGNSGPNYTVGSPAAASKVITVGAVDSDNINASFSSRGPTADGRLKEPVAPGV 352
QY 353 IIAPRASGSMGTPINDYTTKASGTMATPHVSGVALLILOAHPSTPDKVTALLETAD 412
Db 353 IIAPRASGSMGTPINDYTTKASGTMATPHVSGVALLILOAHPSTPDKVTALLETAD 412
QY 413 IYAPKEIADIAVAGAVNNYKAIKYDDYAKLTFTGSVADKGSATHTFVSGATFVATLY 472
Db 413 IYAPKEIADIAVAGAVNNYKAIKYDDYAKLTFTGSVADKGSATHTFVSGATFVATLY 472
QY 473 WDTGSSDIDLXYDPNGNEVDYSYTAYYGFKEKGYNNPTAGTWTVKVSYKGAANYQVDY 532
Db 473 WDTGSSDIDLXYDPNGNEVDYSYTAYYGFKEKGYNNPTAGTWTVKVSYKGAANYQVDY 532
QY 533 VSDGSLSSGCGGNPNPNPPTTDTQTFGTSVNDYMDTSDPTMANVSGATKITGDL 592
Db 533 VSDGSLSSGCGGNPNPNPPTTDTQTFGTSVNDYMDTSDPTMANVSGATKITGDL 592
QY 593 TPDTSYNDIDLXYDPNGNLVDRSTSSNSYEHVEYANPAGTWTFVVAYSTYGMADYOL 652
Db 593 TPDTSYNDIDLXYDPNGNLVDRSTSSNSYEHVEYANPAGTWTFVVAYSTYGMADYOL 652
QY 653 KAVVYVG 659
Db 653 KAVVYVG 659

```

## RESULT 7

US-09-841-553-35

Sequence 35, Application US/09841553  
Publication No. US20020086402A1

## GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru

MORISHITA, Mio

MITTA, Masanori

ASADA, Kiyozo

TSUNASAWA, Susumu

KATO, Ikunoshin

TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSER: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300

CITY: Washington

STATE: D.C.

COUNTRY: United States of America

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/841,553

FILING DATE: 24-Apr-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/894,818

FILING DATE: &lt;Unknown&gt;

APPLICATION NUMBER: JP 323285/1995

FILING DATE: 12-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Browdy, Roger L.

REGISTRATION NUMBER: 25,618

```

REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 654 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-09-841-553-35

```

```

Query Match 72.3%; Score 2483.5; DB 3; Length 654;
Best Local Similarity 72.1%; Pred No. 3.2e-172;
Matches 461; Conservative 69; Mismatches 96; Indels 21; Gaps 8;

QY 1 MKRLGAVLALVVLGSLAAGTALAAPYKPV--VRNNAVQOKNTGLTPGLFKKVMNMNQ 58
Db 1 MKGLKALILVILVLGLVGSVAAPKVEQVRN--VEKNYGLTPGLFKRIKLNPNNE 57
QY 59 EVDYVIMGSGYGRDRAYKVLMLGAQVKYKTIIPAVAVKIKARULLIAGMIDGYRG 118
Db 58 EISTVIVFENHREKEIAVLEIMGAKYVYVHIIPAIADIKVRLDLVIGL--TG--G 113
QY 119 NTRVSGIKFIOEDYKQVQVDA-----TSVSOIGADTVNNSLGVDGCVVAIVDTGIDAN 173
Db 114 KAKLSGVRFIQEDYKTVSAELLEGDESAQWATYVNN--LGDSGGLITGIIDTIDMS 172
QY 174 HPDLKGVIGWYDAVNGRSTPYDDQGHGTHVAGIVAGTSV--NSQYIGVAPGAKLVGVK 232
Db 173 HPDLQKVLGWVDFVNGRSYPYDDHGHGTHVASIAAGTGAASNGKKGAPAKLAGIKV 232
QY 233 LGADSGSVSTIIAGVWVQNKDKYKIRYINLSLSSOSSDCTDLSQAVNNAMAGIV 292
Db 233 LGADSGSVSTIIAGVWVQNKDKYKIRYINLSLSSOSSDCTDLSQAVNNAMAGIV 292
QY 293 VCVAAAGNSGPNYTVGSPAAASKVITVGAVDSDNINASFSSRGPTADGRLKEPVAPGV 352
Db 293 VCVAAAGNSGPNYTVGSPAAASKVITVGAVDSDNINASFSSRGPTADGRLKEPVAPGV 352
QY 353 IIAPRASGSMGTPINDYTTKASGTMATPHVSGVALLILOAHPSTPDKVTALLETAD 412
Db 353 IIAPRASGSMGTPINDYTTKASGTMATPHVSGVALLILOAHPSTPDKVTALLETAD 412
QY 413 IYAPKEIADIAVAGAVNNYKAIKYDDYAKLTFTGSVADKGSATHTFVSGATFVATLY 472
Db 413 IYAPKEIADIAVAGAVNNYKAIKYDDYAKLTFTGSVADKGSATHTFVSGATFVATLY 472
QY 473 WDTGSSDIDLXYDPNGNEVDYSYTAYYGFKEKGYNNPTAGTWTVKVSYKGAANYQVDY 532
Db 473 WDTGSSDIDLXYDPNGNEVDYSYTAYYGFKEKGYNNPTAGTWTVKVSYKGAANYQVDY 532
QY 533 VSDGSLSSGCGGNPNPNPPTTDTQTFGTSVNDYMDTSDPTMANVSGATKITGDL 592
Db 533 VSDGSLSSGCGGNPNPNPPTTDTQTFGTSVNDYMDTSDPTMANVSGATKITGDL 592
QY 593 TPDTSYNDIDLXYDPNGNLVDRSTSSNSYEHVEYANPAGTWTFVVAYSTYGMADYOL 652
Db 588 VPDTSYNDIDLXYDPNGNLVDRSTSSNSYEHVEYANPAGTWTFVVAYSTYGMADYOL 647
QY 653 KAVVYVG 659
Db 648 TAKVYVG 654

```

## RESULT 8

US-10-090-624-16

Sequence 16, Application US/10090624

Publication No. US20020132335A1

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru

```

/ APPLICANT: MORISHITA, Mio
/ APPLICANT: SHIMOJO, Tomoko
/ APPLICANT: ASADA, Kiyozo
/ APPLICANT: KATO, Ikunoshin
/ TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
/ FILE REFERENCE: TAKAKURA-6
/ CURRENT APPLICATION NUMBER: US/10/090,624
/ PRIOR FILING DATE: 2002-03-06
/ PRIOR APPLICATION NUMBER: 09/445,472
/ PRIOR FILING DATE: 1999-12-06
/ PRIOR APPLICATION NUMBER: 151969/1997
/ PRIOR FILING DATE: 1997-06-10
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 16
/ LENGTH: 654
/ TYPE: PRT
/ ORGANISM: Pyrococcus furiosus
US-10-090-624-16

```

```

Query Match 72.3%; Score 2483.5; DB 4; Length 654;
Best Local Similarity 72.1%; Pred. No. 3.2e-172;
Matches 481; Conservative 69; Mismatches 96; Indels 21; Gaps 8;

```

```

QY 1 MKRLGAVLALVVLGLAGTALAAPVKPV--VRNNAVQCKNYGLTPGLFKYQRMNMNQ 58
DB 1 MKGKALLVILVILVGLVGVSAAPBEKVEQVRN---VEKNYGLTPGLFKYQKLNPN 57
QY 59 EVDVTVMFGSYGDRDAVAVLRLMGAQVYKSYKTIIPAVAVKIKARLLLIAGMIDTGYFG 118
DB 58 EISTVIVFNHREKEIAVAVLELMGAKRYVYHIIIPAIADLKVRLDLVLSGL--TG--G 113
QY 119 NTRVSGIKFIEDYKQVODA-----TSVQIGADTVNMSLGVDGSGVVAIVDTGIDAN 173
DB 114 KAKLSGVAFIODEYKVTYSAELGDESAQVMATYVMN--LGVDGSGITIGIIDTGIDAS 172
QY 174 HPDLKGVIGWYDAVNGSRSTPYDDQGHGTHVAGIYAGTGSV--NSQYTGVAAPAKLVGVKV 232
DB 173 HPDLQGVIGWYDAVNGSRSTPYDDHGHGTHVASIAAGTGAASNGKYGMAFGAKLAGIKV 232
QY 233 LGADGSGSVSTIIAGVDVAVVONKDKYGRVNLISGSSQSGDGTDSLQAVNNAMDAQIV 292
DB 233 LGADGSGSVSTIIIGVEMAVDNKDKYGRVNLISGSSQSGDGTDSLQAVNNAMDAQIV 292
QY 293 VCVAAAGNSGPNYTVGSPAAASKVITVGAVDSDNINIASFSSRGPTADGRLKPEVVAQVD 352
DB 293 VVVAAGNSGPNKYTIGSPAAASKVITVGAVDKYDVTISFSSRGPTADGRLKPEVVAQGN 352
QY 353 IIAPRASGSMGTPINDYTTKASGSMATPHVSGVGLIIQAHSPWTPDKVTALILETAD 412
DB 353 IIAPRASGSMGTPINDYTTKASGSMATPHVAGIALLQAHSPWTPDKVTALILETAD 412
QY 413 IYAPKEIADIAGAGRVNVAIKYDDYAKLFTGVSADKGSATHTPDVSGATFVTATLY 472
DB 413 IYKPEIADIAGAGRVNVAIKYDDYAKLFTGYVANKSGQTHQFVISGASFVTATLY 472
QY 473 WDTGSSPIDLYLDPNNGEVDSYTAAYGFEKGYGNPTAGTWTVKVSYKGAANYQVDV 532
DB 473 WDNANSDDLXYLDPNNGQVDYSYTAAYGFEKGYGNPTAGTWTIKVSYSGSANYQVDV 532
QY 533 VSDGSLSGSGGNPNPNPNPTTTDTOTFTGTSVNDYMDTSDPTMNVNSGATKITGDL 592
DB 533 VSDGSLSG-----FGSSPSQPEPTVDAKTFQSGSDHYTDRSDPTMTVNSGATKITGDL 587
QY 593 TPTSTYNDLXYLDPNNGNLVDRSTGSNSYEHVEYANPAPGTWTFVYVASTYGMADYQL 652
DB 588 VPTSTYNDLXYLDPNNGNLVDRSESPNSYEHVEYLPAPGTWTFVYVASTYGMAYTEL 647
QY 653 KAVVYVG 659
DB 648 TAKVYVG 654

```

```

RESULT 9
US-10-800-684-35
/ Sequence 35, Application US/10800684
/ Publication No. US20050014221A1
/ GENERAL INFORMATION:
/ APPLICANT: TAKAKURA, Hikaru
/ APPLICANT: MORISHITA, Mio
/ APPLICANT: YAMAMOTO, Katsuhiko
/ APPLICANT: MITTA, Masanori
/ APPLICANT: ASADA, Kiyozo
/ APPLICANT: TSUNASAWA, Susumu
/ APPLICANT: KATO, Ikunoshin
/ TITLE OF INVENTION: ULTRATHERMOSTABLE PROTEASE GENES
/ FILE REFERENCE: TAKAKURA-1A
/ CURRENT APPLICATION NUMBER: US/10/800,684
/ PRIOR FILING DATE: 2004-03-16
/ PRIOR APPLICATION NUMBER: US/09/841,553
/ PRIOR FILING DATE: 2001-04-24
/ PRIOR APPLICATION NUMBER: 08/894,818
/ PRIOR FILING DATE: 1997-08-29
/ PRIOR APPLICATION NUMBER: JP2385/1995
/ PRIOR FILING DATE: 1995-12-12
/ PRIOR APPLICATION NUMBER: JP96/03253
/ PRIOR FILING DATE: 1996-11-07
/ NUMBER OF SEQ ID NOS: 45
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 35
/ LENGTH: 654
/ TYPE: PRT
/ ORGANISM: Pyrococcus furiosus
US-10-800-684-35

```

```

Query Match 72.3%; Score 2483.5; DB 5; Length 654;
Best Local Similarity 72.1%; Pred. No. 3.2e-172;
Matches 481; Conservative 69; Mismatches 96; Indels 21; Gaps 8;

```

```

QY 1 MKRLGAVLALVVLGLAGTALAAPVKPV--VRNNAVQCKNYGLTPGLFKYQRMNMNQ 58
DB 1 MKGKALLVILVILVGLVGVSAAPBEKVEQVRN---VEKNYGLTPGLFKYQKLNPN 57
QY 59 EVDVTVMFGSYGDRDAVAVLRLMGAQVYKSYKTIIPAVAVKIKARLLLIAGMIDTGYFG 118
DB 58 EISTVIVFNHREKEIAVAVLELMGAKRYVYHIIIPAIADLKVRLDLVLSGL--TG--G 113
QY 119 NTRVSGIKFIEDYKQVODA-----TSVQIGADTVNMSLGVDGSGVVAIVDTGIDAN 173
DB 114 KAKLSGVAFIODEYKVTYSAELGDESAQVMATYVMN--LGVDGSGITIGIIDTGIDAS 172
QY 174 HPDLKGVIGWYDAVNGSRSTPYDDQGHGTHVAGIYAGTGSV--NSQYTGVAAPAKLVGVKV 232
DB 173 HPDLQGVIGWYDAVNGSRSTPYDDHGHGTHVASIAAGTGAASNGKYGMAFGAKLAGIKV 232
QY 233 LGADGSGSVSTIIAGVDVAVVONKDKYGRVNLISGSSQSGDGTDSLQAVNNAMDAQIV 292
DB 233 LGADGSGSVSTIIIGVEMAVDNKDKYGRVNLISGSSQSGDGTDSLQAVNNAMDAQIV 292
QY 293 VCVAAAGNSGPNYTVGSPAAASKVITVGAVDSDNINIASFSSRGPTADGRLKPEVVAQVD 352
DB 293 VVVAAGNSGPNKYTIGSPAAASKVITVGAVDKYDVTISFSSRGPTADGRLKPEVVAQGN 352
QY 353 IIAPRASGSMGTPINDYTTKASGSMATPHVSGVGLIIQAHSPWTPDKVTALILETAD 412
DB 353 IIAPRASGSMGTPINDYTTKASGSMATPHVAGIALLQAHSPWTPDKVTALILETAD 412
QY 413 IYAPKEIADIAGAGRVNVAIKYDDYAKLFTGVSADKGSATHTPDVSGATFVTATLY 472
DB 413 IYKPEIADIAGAGRVNVAIKYDDYAKLFTGYVANKSGQTHQFVISGASFVTATLY 472
QY 473 WDTGSSPIDLYLDPNNGEVDSYTAAYGFEKGYGNPTAGTWTVKVSYKGAANYQVDV 532
DB 473 WDNANSDDLXYLDPNNGQVDYSYTAAYGFEKGYGNPTAGTWTIKVSYSGSANYQVDV 532
QY 533 VSDGSLSGSGGNPNPNPNPTTTDTOTFTGTSVNDYMDTSDPTMNVNSGATKITGDL 592

```

| Db  | 533 | VSDGSLQ-----PSSSPSPQEPFVADKATFGSGSHYYIDSDPTFTVNSGATKLTIDL           | 587 |  |
|---|-----|---|-----|--|
| Qy  | 593 | TFTDSYNDLIDLXYDPNGNLVDIRSTSSSYEHVEYANPAPGTWTFELYAYSTYGMADYQL        | 652 |  |
| Db  | 588 | VFDTSYHDLIDLXYDPNQKLVDRSSSPMSYEHVEYLTAPGWTWELYVAAYTYGMAAYEL         | 647 |  |
| Qy  | 653 | KAVVYTG 659   |     |  |
| Db  | 648 | TAKVYTG 654   |     |  |
| RESULT 10   |     |   |     |  |
| US-10-888-588-16  |     |   |     |  |
| Sequence 16, Application US/10888588                            |     |   |     |  |
| Publication No. US20050084934A1                                 |     |   |     |  |
| GENERAL INFORMATION:  |     |   |     |  |
| APPLICANT: TAKAKURA, Hi karu                                    |     |   |     |  |
| APPLICANT: MORISHITA, Mio                                       |     |   |     |  |
| APPLICANT: SHIMODO, Tomoko                                      |     |   |     |  |
| APPLICANT: ASADA, Kiyoza  |     |   |     |  |
| APPLICANT: KATO, Ikunoshin                                      |     |   |     |  |
| TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE     |     |   |     |  |
| FILE REFERENCE: TAKAKURA=6                                      |     |   |     |  |
| CURRENT APPLICATION NUMBER: US/10/888, 588                      |     |   |     |  |
| CURRENT FILING DATE: 2004-07-12                                 |     |   |     |  |
| PRIOR APPLICATION NUMBER: US/10/090, 624                        |     |   |     |  |
| PRIOR FILING DATE: 2002-03-06                                   |     |   |     |  |
| PRIOR APPLICATION NUMBER: 09/445, 472                           |     |   |     |  |
| PRIOR FILING DATE: 1999-12-06                                   |     |   |     |  |
| PRIOR APPLICATION NUMBER: 151969/1997                           |     |   |     |  |
| PRIOR FILING DATE: 1997-06-10                                   |     |   |     |  |
| NUMBER OF SEQ ID NOS: 33  |     |   |     |  |
| SOFTWARE: PatentIn version 3.0                                  |     |   |     |  |
| SEQ ID NO 16  |     |   |     |  |
| LENGTH: 654   |     |   |     |  |
| TYPE: PRT   |     |   |     |  |
| ORGANISM: Pyrococcus furiosus                                   |     |   |     |  |
| US-10-888-588-16  |     |   |     |  |
| Query Match: 72.3%; Score 2483.5; DV: 5; Length 654;            |     |   |     |  |
| Best Local Similarity 72.1%; Pred. No. 3.2e-172;                |     |   |     |  |
| Matches 481; Conservative 69; Mismatches 96; Indels 21; Gaps 8; |     |   |     |  |
| Qy  | 1   | MKRLGAVLIALVLVLGLAGTALAAFPKPV--VRNNAVOOKYGLTTPGLFKVKORMMNQ 58       |     |  |
| Db  | 1   | MKGLKALLIVLIVLGLGVGVAAPPEKVEQVRN--VEKRYGLTTPGLFKKIQLTANNE 57        |     |  |
| Qy  | 59  | EVDYVIMGSGYGDRAVAVKVLRLMGAOYKYSYKIIIPAVAVKIKARDLLIAGMIDTYGFG 118    |     |  |
| Db  | 58  | EISFYIVPENRREKEIAVRVLELMGAKRYVYHIIIPALIAIDLKVRDLIVISGL--TG--G 113   |     |  |
| Qy  | 119 | NTRVSGIKFIQEDYKVOYVDDA-----TSVSQIGADTVNNSLGIYDSSGVVAIVDTGIDAN 173   |     |  |
| Db  | 114 | KAKLSGARFIQEDYKVTYSABELEGLIDESAQVMATYVNV--LGYDSSGITIGIIDGIDAS 172   |     |  |
| Qy  | 174 | HPDLKGVKIGYVDAVNGRSTPYDDQGHETHVAGIYAGTGSV--NSQYIGVAPGAKLVGYKV 232   |     |  |
| Db  | 173 | HPDLQGVKIGYVDPFNGRSYPIYDDHGHGTHVASTIAGTGAASNGRYKMGAPGAKIAGIKV 232   |     |  |
| Qy  | 233 | LGADSGGSVSTTIAGVDVWVONKDKYGIKRVINLSIGSSQSDGTDLSLQAVNNAMADGIV 292    |     |  |
| Db  | 233 | LGADSGGSISITIKIGVEMAVDNKDKYGIKVINLSIGSSQSDGTDLSQAVNAWDAGLV 292      |     |  |
| Qy  | 293 | VCVAAGNSGPNITYTIGSPAAASKVITTVGAVDSNDNIAFPSSRSGPRAOGRLEKREVAAPGV 352 |     |  |
| Db  | 293 | VVVAAGNSGPNKTYTIGSPAAASKVITTVGAVDXYDTITSFSSRGPRAOGRLEKREVAAPGNW 352 |     |  |
| Qy  | 353 | IIPAPASGTSMGTPINDYYTKRSGTSMATPHNSGVALIILQHPSPMTEDKVTALIERAD 412     |     |  |
| Db  | 353 | IIPAPASGTSMGCPINDYYTAAFGTSMATPHVAGIALLILQHPSPMTEDKVTALIERAD 412     |     |  |
| Qy  | 413 | IVAPKEIADIVAGRAVNVKAIKYDYAKLTFTGSVADKGSANTHTPEVSGATFEVATLY 472      |     |  |

[illegible]

|                       |                  |                     |           |             |
|-----------------------|------------------|---------------------|-----------|-------------|
| Query Match           | 62.2%            | Score 2138.5;       | DB 3;     | Length 522; |
| Best Local Similarity | 78.3%            | Pred. No. 3.4e-147; |           |             |
| Matches 407;          | Conservative 36; | Mismatches 70;      | Indels 7; | Gaps 3;     |

|    |     |  |     |
|----|-----|--|-----|
| Qy | 141 | SVSGIGADPTWNS:LGYYGSGSVVAIV:YDTGIDAHNPL:KCKYV:GWDAN:NGSGSTPDQGH        | 200 |
| Db | 9   | SAQVAMTYWN-LGYDGSGLTIGITDGTGDASHPLQSKYV:GAWDFVNGSGSYFDHGH              | 67  |
| Qy | 201 | GTHVAGVYAGTGSV-NSQY:GVAPGAKL:GVGVKYL:GADSGSVSTI:IAGDWVYONKRYG          | 255 |
| Db | 68  | GTHVAS:IAAGGAASNCKYKGMAPGAKL:GIKVL:GADGSGSISTII:IKGEVANAONKRYG         | 12  |
| Qy | 260 | IRVNTLS:GSSGSSDGTDSL:GSAVNNAMAGITV:CVAA:NSGPNMTYTGSGPAAASKYITV         | 315 |
| Db | 128 | IKVNTLS:GSSGSSDGTDSL:GSAVNNAMAGL:VYVVA:NA:SGPNKRTTIGSPAAASKYITV        | 187 |
| Qy | 320 | GAVDSNDN:IASPSSRGFTADGRL:KEPVA:PGVDII:APRAGTSMGT:INDYYT:YASGTSM        | 379 |
| Db | 188 | GAVDKXVIVITSFSSRGFTADGRL:KEPVA:PGWMI:IAARASGTSMGQ:INDYYT:AAPGTSM       | 247 |
| Qy | 380 | ATPVS:GVGL:II:OAHPSMT:PDVKYKAL:ETAD:YAP:KE:AD:IAN:GAARVNVY:QAL:KYD     | 439 |
| Db | 248 | ATPVS:AG:IAAL:II:OAHPSMT:PDVKYKAL:ETAD:YAP:KE:AD:IA:GAARVANA:YIA:IN:YD | 307 |
| Qy | 440 | YAKLFTG:SVADK:GSA:HTFDVSGAT:FTVATL:YMDTGS:SDIDL:YLYDNG:ENVDYS:TA       | 499 |
| Db | 308 | YAKLFTGVY:VANKGS:QTHQFV:ISGASFVATL:YMN:NASD:IDL:YLYDNG:QVDSY:TA        | 367 |
| Qy | 500 | YGEFKYGVY:NP:PLAGT:TVK:VY:SYKGAAN:QVUDV:VDSGL:SGSGG:GNPN:PNPNPT:TTD    | 559 |
| Db | 368 | YGEFKYGVY:NP:PLDGT:TVK:VY:SYKSGGAN:QVUDV:VDSGL:SQ-----PSSPS:PP:BE:TTD  | 422 |
| Qy | 560 | TQTFG:SVNDY:MDTSDTFT:MMNVNSGAT:IKT:GDL:TEDT:SYNDL:DL:YLYD:PNCL:VDR:TS  | 619 |
| Db | 423 | AKTFQXSDHY:YDRSDTFT:MTVNSGAT:IKT:GDLVFDTS:YHD:DL:YLYD:PNQKLYDR:BS      | 482 |
| Qy | 620 | NSYEH:EVAN:PA:GTT:FLV:YAYSTY:GMA:DV:Q:LA:V:VYTG                        | 659 |
| Db | 483 | NSYEH:EVYL:PA:GTY:FLV:YAYY:Y:GMA:YEL:TK:VYTG                           | 522 |

RESULT 12

```

1      / Sequence 4, Application US/10090624
2      / Publication No. US2002013335A1
3      / GENERAL INFORMATION:
4      / APPLICANT: TAKAKURA, Hikaru
5      / APPLICANT: MORISHITA, Mio
6      / APPLICANT: SHIMOTO, Tomoko
7      / APPLICANT: ASADA, Kiyozo
8      / APPLICANT: KATO, Ikunoshih
9      / TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
10     / FILE REFERENCE: TAKAKURA-6
11     / CURRENT APPLICATION NUMBER: US/10/090,624
12     / CURRENT FILING DATE: 2002-03-06
13     / PRIOR APPLICATION NUMBER: 09/445,472
14     / PRIOR FILING DATE: 1999-12-06
15     / PRIOR APPLICATION NUMBER: 151969/1997
16     / PRIOR FILING DATE: 1997-06-10
17     / NUMBER OF SEQ ID NOS: 33
18     / SOFTWARE: PatentIn version 3.0
19     / SEQ ID NO 4
20     / LENGTH: 522
21     / TYPE: PR1
22     / ORGANISM: Pyrococcus furiosus
23     / FEATURE:
24     / NAME/KEY: misc_feature
25     / LOCATION: (428)..(428)
26     / OTHER INFORMATION: xaa at position 428 is Gly or Val.
27     / US-10-090-624-4

```

Best Local Similarity 78.3%; Pred. No. 3.4e-147;  
Matches 407; Conservative 36; Mismatches 70; Indels 7; Gaps 3;

```

QY 14 SVS0IGADDTVMNLSIGYDGSQVVAIYVDGIDIANHFDLKGXIGWDAVANGSTYDDQGH 200
Db 9 SAAQWALTYVMN-LGTDSSGRTTIGIITDIDISHPDLGKXIGWDFPNANGSTYDDHGH 67
QY 201 GTHVAGIYVAGTGSV-NSOYIGVAPGAKLVGVKVLGADSGSVSTIIAGVDWVONKDKYG 255
Db 68 GTHVASIAGDTGAASNGKXKGMAPGAKLAGIVLGADSSGSIITIKGVEWAVDNKDKYG 127
QY 260 IRVYNLSIGSSQSSDGTDSLQAVNNAMNADAGIYVCAAGNSGPNPTYTWGSPAASKYITV 319
Db 128 IKYNLSLSSQSSDGTDLQAVNANAMNADAGIYVVAAGNSGPNKXITIGSPAASKYITV 187
QY 320 GAVDSNDNIASFSSRGPJADGRLKEVEVAPGVDIIAAPRASGTSMGTPINDYTTKASGTM 379
Db 188 GAVDKYDVLTSSFSRRPJPADGRLKEVEVAPGWNIIIAAASGTSMGQFINDYTTAAPGTM 247
QY 380 ATPHVSQVGLIIIOAHPSWTPDKXVTALIEPADIVAPKEIDIVYAGRNVVYAKICYD 439
Db 248 ATPHVAGIATLLOAHPSWTPDKXVTALIEPADIVKPELADIVYAGRNVVYAAINYN 307
QY 440 YAKLFTGSVADKGSATHTFDVSGATFYATLTYMDTSSDIDLXYPNGENVDYSTAY 499
Db 308 YAKLFTGVYVANKGSQTHQFVIGSAPFYATLTYMDNANSDDLXYDPNGQVDYSTAY 367
QY 500 YGEXEKGYNPNPAGTMYVVSYKKAANYQVDVYSDGSLSSGGGANPNPNPNPPTTTD 559
Db 368 YGEXEKGYNPDGWTTLIKVVSYSANAYQVDVYSDGSLD-----PGSSPSPQPEPTVD 422
QY 560 TOTFTGSVNDYMDSDTFTTMVNSGATKIITDILFTDSYNDLJLYLDPNGNIVDRSTSG 619
Db 423 AKTFQXSDHYIYDRSDTFTTMVNSGATKIITDILVFDITSYHDLJLYLDPNGKLVDSBSB 482
QY 620 NSTEHEFYANPAPGTMTPLYVAYSTYGMADYQKAAVYTG 659
Db 483 NSTEHEVYLTLPAGTMTPLYVAYSTYGAAYELTAKYTYG 522

```

### RESULT 13

```

US-10-800-684-3
/ Sequence 3, Application US/10800684
/ Publication No. US20050014221A1
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hi-karu
APPLICANT: MORISHITA, Mitsu
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Kuniooshin
FILE REFERENCE: ULTRATHERMOSTABLE PROTEASE GENES
CURRENT APPLICATION NUMBER: US/10/800,684
CURRENT FILING DATE: 2004-03-16
PRIOR APPLICATION NUMBER: US/09/841,553
PRIOR FILING DATE: 2001-04-24
PRIOR APPLICATION NUMBER: 08/894,818
PRIOR FILING DATE: 1997-08-29
PRIOR APPLICATION NUMBER: JP92385/1995
PRIOR FILING DATE: 1995-12-13
PRIOR APPLICATION NUMBER: JP96/03253
PRIOR FILING DATE: 1996-11-07
NUMBER OF SEQ ID NOS: 45
SOFTWARE: Seqin version 3.2
/ SEQ ID NO 3
/ LENGTH: 522
/ TYPE: prt
/ ORGANISM: Pyrococcus furiosus
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (428)..(428)

```



QY 141 SVEQIGADTVWNSIGYDGSVVVAIVDTGIDANHPDLKGKVI GWYDAVNGRSTPYDPQGH 200  
 Db 9 SAAQVMAITYWMN-LGYGSGITIGIIDTGIDASHPDLQKVI GWYDAVNGRSTPYDDHGH 67  
 QY 201 GTHVAGIVAGTGSV-NSQYIGVAPGAKLVGVKVI GADGSGSVSTIIAGVDMVVQNKDKYK 259  
 Db 68 GTHVASIAGTGAASNGKYGMAFGAKIAGIKVLGADGSGSISTIIKGVEMAVDNKDKYK 127  
 QY 260 IRVINISLGSSQSGDGTDSLQAVNNMADAGIVVCVAGNSGPNYTVGSPAAASKVITY 319  
 Db 128 IKVINISLGSSQSGDGTDSLQAVNANMADAGLVVVAAGNSGPNKYTIGSPAAASKVITY 187  
 QY 320 GAVDSNDNIAFSRSGPTADGRLKPEVAPGVDIIPRASGSMGTEINDYTKASGTSM 379  
 Db 188 GAVDKYDVITSFSRSGPTADGRLKPEVAPGVNMIARASGSMGPIINDYTTAPGTSM 247  
 QY 380 ATPHVSQVGLIIOAPSWTPDKVKTLIETADIAPKEIADIAYGAGRVNYYKAIKYDD 439  
 Db 248 ATPHVAGIAALLQAPSWTPDKVKTLIETADIVKPEIADIAYGAGRVNAYKAIINYDN 307  
 QY 440 YAKLTFGSAVDKGSATHTFDVSGATPVTATLYMDTGSDDILYLYDPNGNEVDYSYAY 499  
 Db 308 YAKLVFTGYVANKSQTHTQFVIGASFTVATLYMDNANSDDLXYLYDPNGNOVDYSYAY 367  
 QY 500 YGFEKVGYNPTAGTWTVKVVSYKGAANYQVDVVSQSLSQSG 542  
 Db 368 YGFEKVGYNPTDGTWTIKVVSYSGSANYQVDVVSQSLSQSG 410

Search completed: January 6, 2007, 22:37:52  
 Job time : 124.5 secs